Development of a Hybrid Model for Freeway Incident Duration: A Case Study in Maryland

Woon Kim
Research Assistant, Civil and Environmental Department
University of Maryland at College Park
1173 G. Martin Hall, Dept. of CEE, University of Maryland at College Park, MD 20740 USA
+1-301-405-2638, wkim2@umd.edu

Gang-Len Chang
Professor, Civil and Environmental Department
University of Maryland at College Park
1173 G. Martin Hall, Dept. of CEE, University of Maryland at College Park, MD 20740 USA
+1-301-405-1953, gang@umd.edu

ABSTRACT

This paper presents a methodology for estimating the incident duration and identifying its critical contributing factors in the state of Maryland. The incident database from Year 2003 to Year 2006 from the Maryland State Highway (MDSHA) and Police Accident Report database were used for model development. This study employed a hybrid model, consisting of a Rule-Based Tree Model (RBTM), Multinomial Logit Model (MNL) and Naïve Bayesian Classifier (NBC) to develop the primary estimation system. Through the model development, we have identified a number of remarkable findings regarding relationships between the set of key factors and incident duration. The proposed model along with findings will play a vital role for traffic agencies to establish an advanced traveler information system that can provide the incident-induced delay to both pre-trip and the en-route drivers.

Key Words: incident duration, tree model, multinomial logit model, Naïve Bayesian Classifier, incident management

INTRODUCTION

Incidents on congested highways, regardless of involving personal fatalities, injuries, or property damages will all cause considerable reduction in capacity, followed by heavy congestion and delay, and thus give birth to the enormous socioeconomic loss (1). With the reliably predicted information of incident duration, responsible agencies can inform travelers with timely updated Variable Message Signs, estimate the resulting queue length, and assess the need to implement detour as well as control operations. Thus, an effective model to predict incident duration can be a valuable tool to mitigate non-recurrent congestion. The objective of this paper is to present a methodology for estimating incident duration and identifying critical variables in the state of Maryland using the MDSHA (the Maryland State Highway) incident database and Police Accident Report collected from Year 2003 to Year 2006. Note that in this study, the estimation of incident duration is based on the range of time interval it may fall, such as between 15~30 minutes due to the preference of incident response operators from the perspective of both the application and the system reliability.

LITERATURE REVIEW

Incident duration has been studied by numerous researchers for several decades with various methodologies. Some of those popular approaches are [1] Probabilistic Distributions (2), [2] Conditional Probabilities (3), [3] Linear Regression Models (4), [4] Time Sequential Models (5) [5] Decision Trees and Classification Trees (6), and [6] Discrete Choice Models (7). In addition, there are some recent research using Artificial Neural Network (8) and information-based time sequential approach (9). Although there are a variety of existing techniques with acceptable results, most research findings are not transferable to other locations. Each model was developed with different incident data sources and descriptive variables, and thus yields somewhat different results. Therefore, for any target application, it is necessary to develop a new model for different traffic environments and available data sources.

DATA NATURE

DATA DESCRIPTION

For this study, highway incident data extracted from CHART-II Database (Coordinated Highways Action Response Team) and Accident Report DB were used. The CHART II database maintained by MDSHA contains information about the details associated with each incident occurring in Maryland, including nature of incident, time of detection/occurrence, response time, vehicles involved, types of involved vehicles, number of lanes/shoulders blocked, responded unit, pavement conditions, etc. On the other hand, the Accident Report DB is the primary source providing more extensive information for severe incidents, including fatalities and personal injuries, such as number of fatalities/injuries, collision types, light conditions, etc.

INCIDENT DURATION NATURE

Before starting the model development, we first investigated the relationship between incident nature and the resulting duration. Figure 1 illustrates the distribution of incident duration frequency and the statistics of four major incident types: collision-property damage, collision-personal injury, disabled vehicles and collision-fatality. As reflected in the figures, the incident duration exhibits remarkably different distributions between different incident types. For instance, incidents involving disabled vehicles and property damage are likely to have a shorter duration, while incidents causing personal injuries are more likely to have a longer duration. Notice that majority of fatality incidents lasted more than 2 hours. In addition, each distribution disperses in a quite wide range without exhibiting any distinctive patterns. Hence, it is unlikely to fit the incident duration data with any particular type of continuous or discrete statistical distribution.

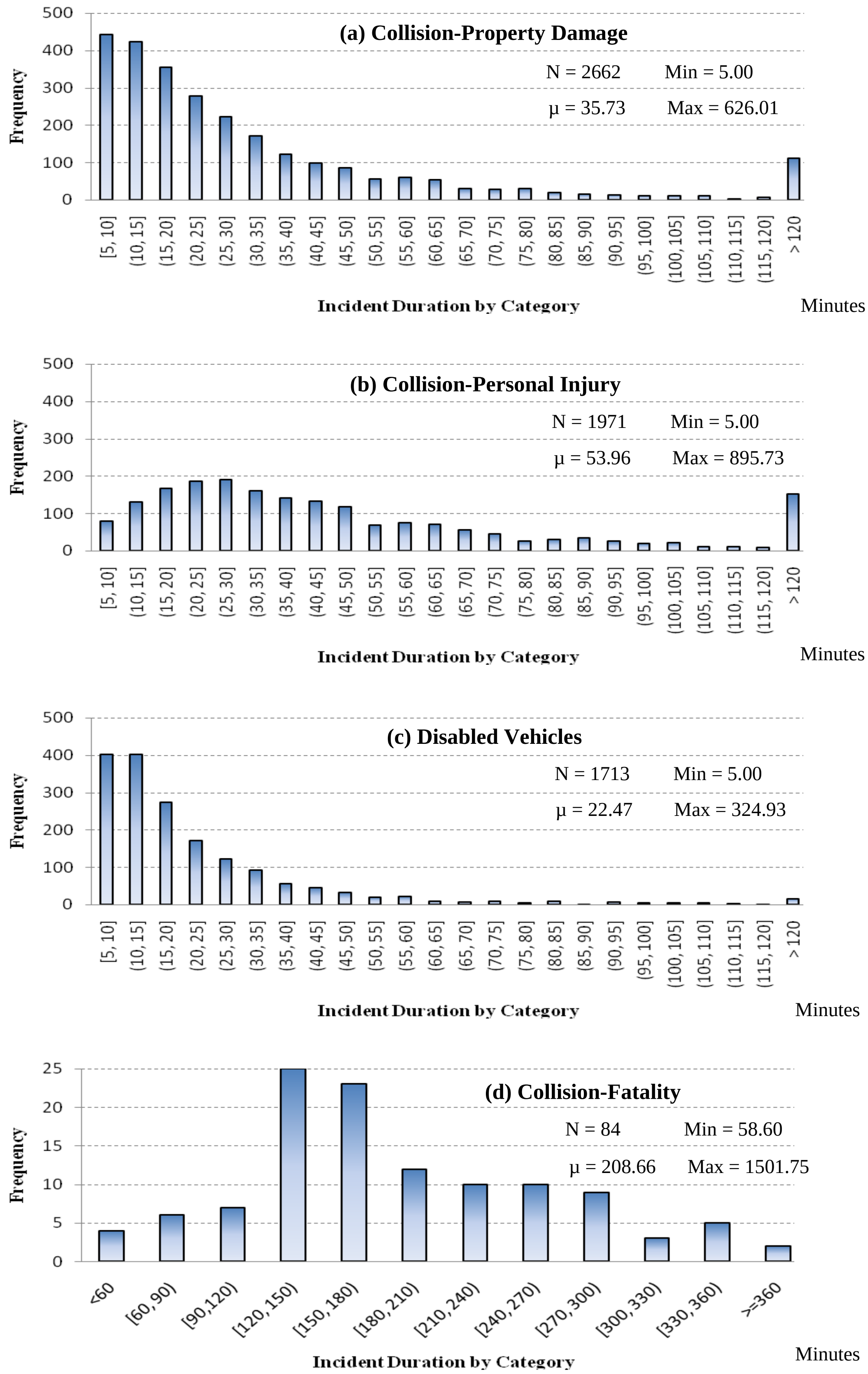


Figure 1. Distributions of Incident Duration by Incident Nature

METHODOLOGY AND ANALYSIS

Based on the results of literature review and the analysis of incident duration nature, we have selected a hybrid model as the primary methodology, which consists of two components: non-fatality and fatality involved incidents. The model for non-fatality involved incidents was developed with the Rule-Based Tree Method (RBTM) as its primary module and the multinomial logit model as a supplemental module, while Naïve Bayesian Classifier (NBC) was used to develop the model for fatality-related incidents.

In this section hereafter, data from incidents involved in personal injury only will be utilized as an example to illustrate the RBTM development procedure, whereas fatality incidents will be employed for the NBC development.

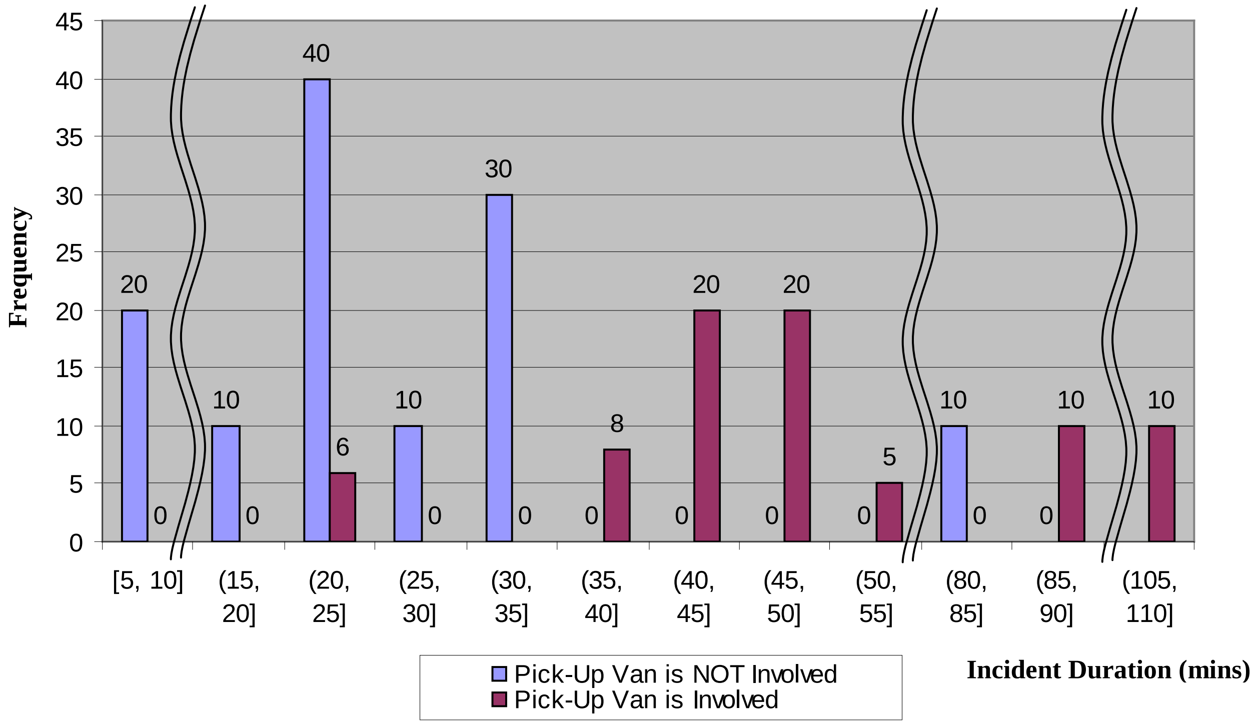
RULE-BASED TREE MODEL (RBTM)

A tree model has long been used for both classification and prediction purpose due to its independence of distributional assumption and the flexibility to fit any discrete data patterns. Based on the findings from our preliminary analysis, this study has redesigned a conventional tree model, named a Rule-Based Tree Model (RBTM), using the following procedures. Note that incident duration data were grouped into 5-minute intervals for model development because of the precision issue associated with data recording and the response operators in managing incidents.

Step 1: <u>Set the incident nature as the first splitter.</u>
In this research, incident nature was categorized into collision-fatality (CF), collision-personal injury (CPI), collision-property damage (CPD), disabled vehicles (Disabled) or others (Others) which include all other kinds of incidents. Due to the relatively small sample size, the category of Others, including fire, road debris, constructions, and police activities, were excluded from this study. Since the incident duration displays obviously different distributions by incident nature as shown in Figure 1, it was selected as the first splitter for developing RBTM.

Step 2: <u>Determine the next splitter for each node.</u>
This step is to generate a crosstabulation table (10) and to determine the next splitter for each node. A crosstabulation table can display the number of cases in each category defined by two or more specified variables. For each independent and dependent variable (i.e., incident duration), this step shall create a crosstabulation table along with a bar chart to show the distribution of frequency for different categories of the independent variable that is potentially associated with the incident duration. Then, the independent variable that can classify the incident duration data into two most distinctly different categories shall be selected as the next splitter.

Figure 2 will provides better understanding of this step. These cases used for this figure is taken from some part of the dataset used for this research. Assume that this bar chart is created based on the incidents whose nature is collision-personal injury from Step 1. As shown in the figure, a frequency bar chart is created for each category of selected independent variable, that is, whether any pick-up van is involved with incidents or not. In this instance, these two categories display obviously different distributions. Incidents without pick-up van involvement are highly likely to be cleared within 30 minutes, while those involved pick-up vans are more likely to last longer. To see if there are any other independent variables that can best classify the available data into different categories, we create this kind of bar charts for every available independent variable. After comparing the results, we select the most critical independent variable for each category as the next splitter.



*Note: some of incident duration ranges are omitted since their frequencies are zero, and they are replaced with a symbol .

Figure 2. Distributions of Categories for a Pick-Up Van Involved Indicator

Step 3: <u>Split the node based on the determined splitter in each category.</u>
The focus of this step is to convert each splitting node into *If-then; Else-then* statement, which will constitute the set of rules for determining the incident duration for the node.

For instance, consider an example presented in the previous steps whose nature is collision-personal injury (CPI). Also, assume that the pick-up van involvement indicator is selected as the next splitter. Then, the split nodes on the basis of categories of this splitter can be presented as *If-then; Else-then* statements such as *If Incident Nature is CPI & Pick-up Van is Not involved, then Incident Duration is ...; If Incident Nature is CPI & Pick-up Van is involved, then Incident Duration is ….*

Step 4: <u>Assign the estimated/predicted incident duration range for each split node.</u>
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 within 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.

In Figure 2, most incidents that did not involve pick-up vans are distributed in the range of 5~30 minutes. Since this incident duration range is less than 30 minutes and covers about 92% (110/120) of all cases within the given conditions (within this node), 5~30 minutes will be the most plausible incident duration interval in this node. On the other hand, incidents involving pick-up vans are more widely distributed, and thus it is more difficult to determine the incident duration range under the given conditions. In this instance, first, we need to find out if any 30 minute-interval can cover the most of incidents, and it turns out to be the interval of 20~50 minutes that contains approximately 68% of incidents in the node. Since it does not exceed 70%, we find any shortest duration interval cover at least 60% of all cases. In the figure above, it appears to be the interval of 35~50 minutes with about 61 percent of coverage.

Step 5: <u>Repeat Step 2 to Step 4 for all nodes until satisfy the predetermined criteria for stopping the tree growth.</u>

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

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

Repeating Step 2 to Step 4 will help improve the model performance. Considering the instance presented previously, adding another splitter can either narrow the range of incident durations estimated/predicted or increase the proportion of incidents covered by that range of incident duration. This procedure can be repeated until no independent variable that can further be used to divide the data in its category into different distributions.

Figure 3 describes the structure of the Rule-Based Tree Model (RBTM). The developed RBTM starts with the first splitter, and the second splitter is determined independently for incidents falling in each category of the first splitter. In Figure 3, categories of the first splitter are expressed with a subscript 1, i.e., *Category* A_1 or *Category* B_1. Note that the second splitters can be different for each subset of incidents, depending on its characteristics and distributions. Figure 3 reflects this feature with numbering the 2nd splitters such as *2nd Splitter-1* or *2nd Splitter-2*. Likewise, in the figure the categories of second splitters are subscribed with numbers, e.g., 2-1 or 2-2, to distinguish multiple splitters at the same level. The tree model continues to extend with third splitters and so on. In this study, the first splitter turns out to be incident nature and it has five categories, which creates five branches extending from the first splitter.

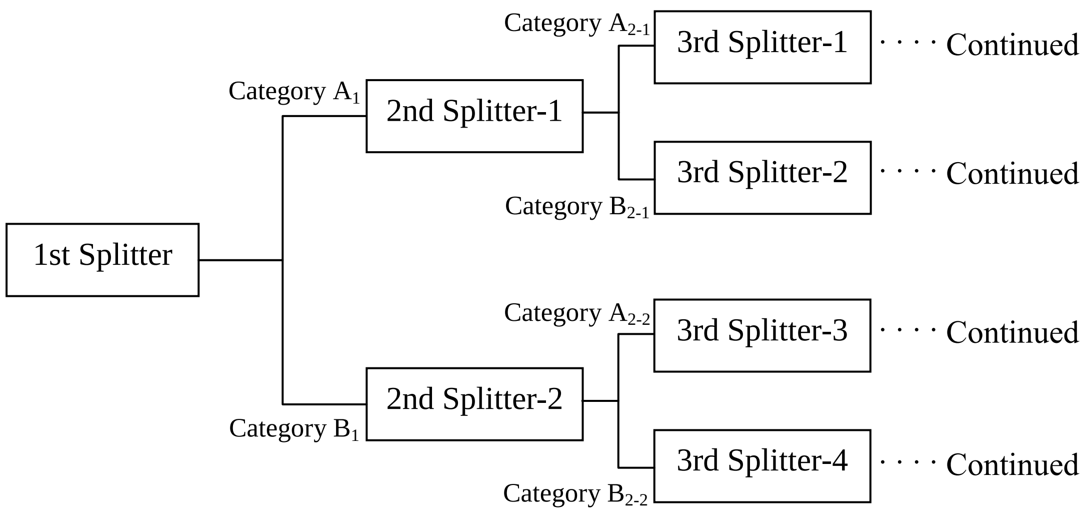


Figure 3. Illustration of the Structure of the Rule-Based Tree Model

MNL SUPPLEMENTAL MODELS

For each of non-fatality involved incidents, we have explored the use of supplemental models to improve the resulting accuracy, since one shall not expect that the Rule-Based Tree Model can reflect all embedded relations and provide the performance sufficiently reliable for real-world application. As a supplemental model, Multinomial Logit (MNL) Models were developed to estimate the relation between incident duration and its associated factors. A well calibrated model will allow its users to predict the duration category a detected incident belongs to, based on the estimated probabilities of all incident duration categories. The core concept of MNL is same as that used in accident severity model (11). A detailed discussion regarding MNL models would be found in the references (12), (13) and (14).

Figure 4 displays the structure of the proposed hybrid model using RBTM and MNL to estimate non-fatality incidents. Table 1 summarizes the calibrated MNL models for collision-personal injury incidents, followed by Table 2 to describe variables included in those models.

NAÏVE BAYESIAN CLASSIFIER FOR NON-FATALITY INCIDENTS

Unlike other incident natures, fatality incidents have relatively small samples with an extremely wide range of duration distribution. Also, the majority of fatality-involved incidents last from two hours up to several hours which are very different from other incident types. These unique features of incidents resulting in fatalities present the Rule-Based Tree Model from providing satisfactory classification. Thus, Naive Bayesian Classifier was selected as an alternative approach to develop a fatality incident duration model. This section illustrates the background of methodology briefly, followed by the structure of the developed model.

Naïve Bayesian Classifier

The NBC assigns the object I to one of the discrete categories, $D_1, D_2, \cdots, D_m$, based on its attributes, $X_1, X_2, \cdots, X_m$. The NBC calculates the probability that I belongs to each category, conditioning on the observed attributes. I is assigned to the category with the greatest such probability. This classifier is based on applying Bayes' theorem with the assumption that the presence of a specific attribute is unrelated to the presence of any other attributes. The probability that I belongs to each category is calculated on the observed attributes, that is, $P(I \in D_i \mid X_1, X_2, \cdots, X_n)$. Applying Bayes' Theorem, this is rewritten as

$$P(I \in D_i \mid X_1, X_2, \cdots, X_n) = \frac{P(I \in D_i)\, P(X_1, X_2, \cdots, X_n \mid I \in D_i)}{P(X_1, X_2, \cdots, X_n)} \qquad \text{(Eq. 1)}$$

Under the mutual conditional independence assumption, this reduces to

$$P(I \in D_i \mid X_1, X_2, \cdots, X_n) = \frac{P(I \in D_i) \prod_{j=1}^{n} P(X_j \mid I \in D_i)}{P(X_1, X_2, \cdots, X_n)} \qquad \text{(Eq. 2)}$$

for each category D_i. Since the denominator will be the same for all categories, we need only calculate the numerator for each category i, choosing

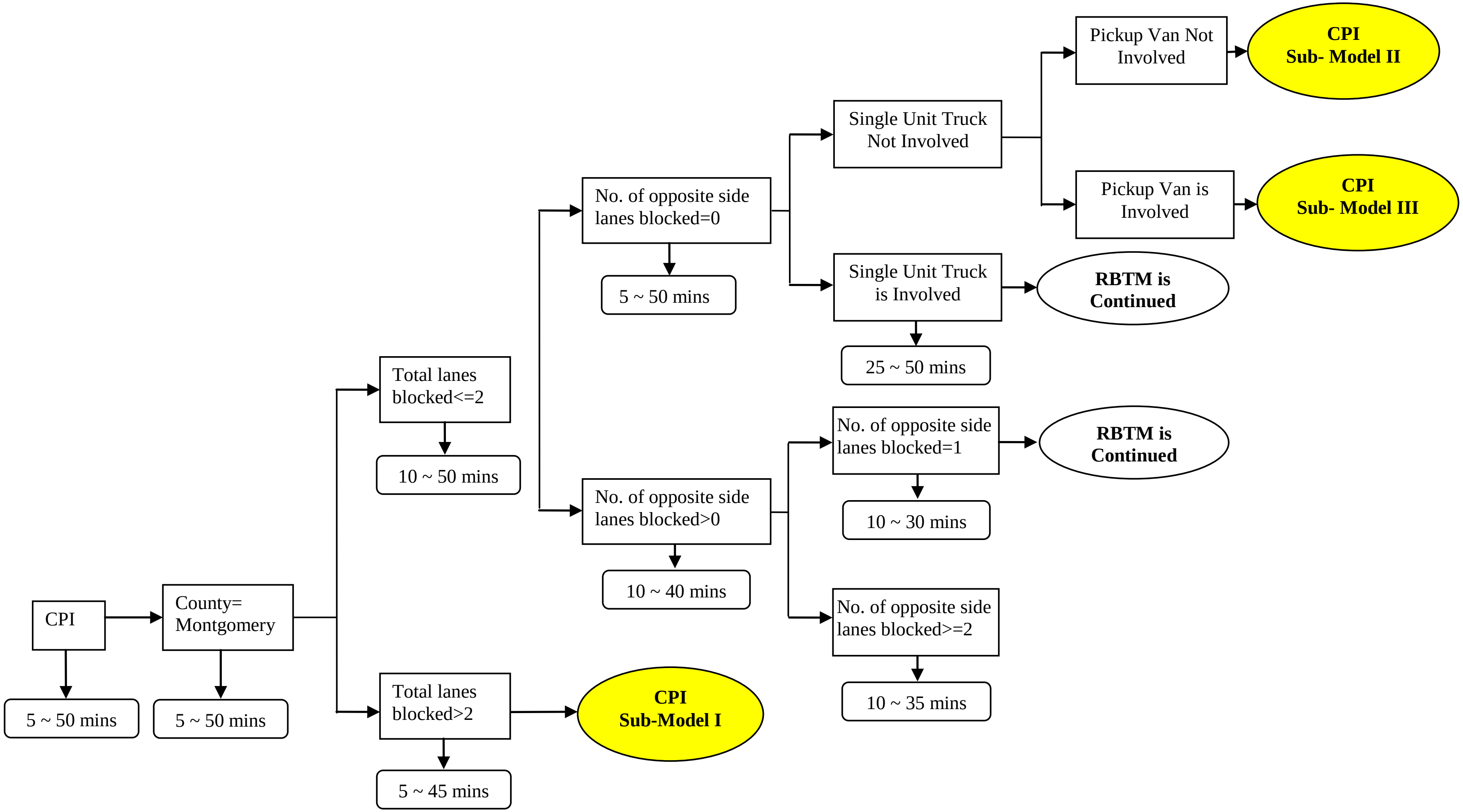


Figure 4. Rule Based Tree Model and Structure of the Proposed Hybrid Model for Collision-Personal Injury

Table 1. Calibrated MNL Models for Collision-Personal Injury

CPI-Sub-Model I

$$R_{5\text{-}25} = \underset{(0.9)}{0.910} \underset{(-2.9)}{-3.550}*\text{NoTT} \underset{(-2.4)}{-2.140}*\text{Night} \underset{(-2.4)}{-0.536}*\text{NoVehInv} \underset{(3.2)}{+2.434}*\text{I495} \underset{(-3.3)}{-3.053}*\text{NoSUT} \underset{(-2.3)}{-0.971}*\text{NoPUV} \underset{(1.6)}{+1.053}*\text{Pave\_Dry}$$

$$R_{25\text{-}45} = \underset{(2.9)}{2.131} \underset{(-2.0)}{-1.241}*\text{NoTT} \underset{(-3.2)}{-2.678}*\text{Night} \underset{(-2.4)}{-0.536}*\text{NoVehInv} \underset{(1.9)}{+1.253}*\text{I495} \underset{(-3.3)}{-3.053}*\text{NoSUT}$$

R_{gt45} = 0 (Base)

The number of observations used : 98
Likelihood with zero coefficients = -106.5654
Likelihood with constants only = -105.5362
Final value of Likelihood = -76.2511

Note : Numbers in parentheses are *t*-statistic values.

CPI-Sub-Model II

$$R_{5\text{-}25} = \underset{(2.5)}{1.952} \underset{(2.0)}{+1.827}*\text{I270} \underset{(-3.1)}{-0.655}*\text{NoVehInv} \underset{(2.3)}{+2.663}*\text{I495} \underset{(-2.7)}{-2.776}*\text{Pave\_SI} \underset{(-2.1)}{-2.050}*\text{Ex495}$$

$$R_{25\text{-}50} = \underset{(2.0)}{1.576} \underset{(1.8)}{+1.568}*\text{I270} \underset{(-2.2)}{-0.422}*\text{NoVehInv} \underset{(2.1)}{+2.471}*\text{I495} \underset{(-2.7)}{-3.626}*\text{Pave\_SI} \underset{(-2.3)}{-2.253}*\text{Ex495}$$

R_{gt50} = 0 (Base)

The number of observations used : 189
Likelihood with zero coefficients = -206.5391
Likelihood with constants only = -179.5752
Final value of Likelihood = -167.4129

Note : Numbers in parentheses are *t*-statistic values.

CPI-Sub-Model III

$$R_{5\text{-}25} = \underset{(2.8)}{1.868} \underset{(-3.2)}{-3.346}*\text{NoTT} \underset{(-2.1)}{-2.773}*\text{Night} \underset{(-2.2)}{-2.509}*\text{PEAKHR} \underset{(-3.6)}{-3.874}*\text{Ex270}$$

$$R_{25\text{-}45} = \underset{(3.8)}{3.031} \underset{(-3.2)}{-3.346}*\text{NoTT} \underset{(-1.7)}{-1.603}*\text{Night} \underset{(-1.9)}{-2.095}*\text{PEAKHR} \underset{(-3.1)}{-2.727}*\text{Ex270} \underset{(-1.5)}{-0.865}*\text{Ex495} \underset{(-2.1)}{-1.099}*\text{Pave\_Dry}$$

R_{gt45} = 0 (Base)

The number of observations used : 82
Likelihood with zero coefficients = -90.0862
Likelihood with constants only = -85.9470
Final value of Likelihood = -65.3223

Note : Numbers in parentheses are *t*-statistic values.

Table 2. Descriptions for Variables Included in the CPI-Sub-Models

Description for Variables Included in the CPI-Sub-Models	
I495	1 if an incident occurred on Road I-495; 0 otherwise
Night	Binary variable for incident time (Night=1, otherwise=0)
NoTT	Number of tractor-trailers involved
NoPUV	Number of pickup/vans involved
NoVehInv	Number of vehicles involved
NoSUT	Number of Single-Unit Truck involved
Pave_Dry	1 if pavement condition is Dry; 0 otherwise
I270	1 if an incident occurred on Road I-270; 0 otherwise
Ex495	Binary variable to indicate the specific locations on I-495 (exit numbers. 27, 28, 33, 34, 36, 38, and 39)
Pave_SI	1 if Pavement Condition is Snow/Ice; 0 otherwise
Ex270	Binary variable to indicate the specific locations on I-270 (exit numbers. 1, 4, 9, 13, 15, 18, and 22)
PEAKHR	PEAKHR : 1 if an incident occurred in peak hours; 0 otherwise

$$i^* \in \arg\max P(I \in D_i) \prod_{j=1}^{n} P(X_j \mid I \in D_i)$$

and assigning I to category D_{i^*}.

Seeing the incident duration prediction, the attributes X correspond to observable incident characteristics, such as pavement conditions, locations of incidents, the number of vehicles involved, the number of blocked lanes and so on. It is also necessary to define discrete categories of incident duration to classify the incidents. When an incident occurs, the NBC would calculate the probability that the incident's duration will fall into each discrete category and decide the incident duration with the highest probability.

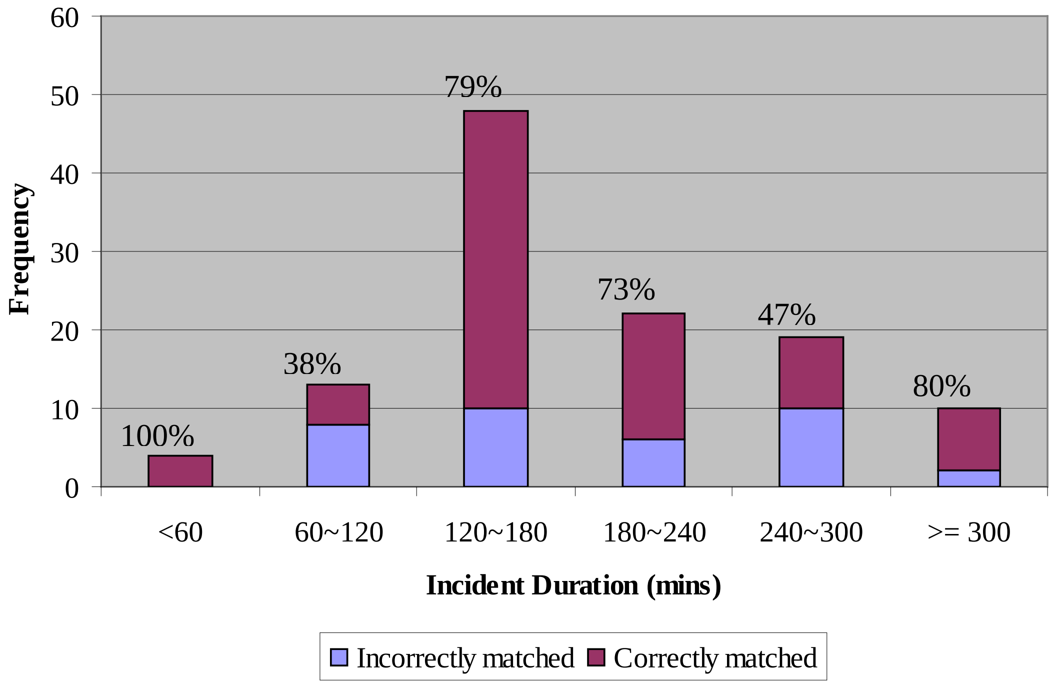
This probability calculation has an advantage that can still compute the probability even if some of the attributes are not present. Regardless of how much or how little is known about the incident, a valid prediction is possible based on the attributes provided. All model parameters (*i.e.*, the probabilities $P(I \in D_i)$ and $P(X_j \mid I \in D_i)$) can be approximated with relative frequencies from the training set. These are maximum likelihood estimates of the probabilities. If the given category and attribute values never occur together in the training set, then the frequency-based probability estimate will be zero. This is problematic since it will wipe out all information in the other probabilities when they are multiplied. For this reason, zero probabilities are replaced by a small positive number when calculating these products.

Model Development and Results

Since not all attributes can yield positive impact for projecting the incident duration, it is necessary to explore which attributes can best improve the NBC model. To do so, we first developed a simple Naïve Bayes Classifier (NBC) model for each attribute only. By comparing the estimating results, one can select attributes with best results as the initial set. For each of those selected attributes, we added another attribute to create a two-attribute set. Then we compared and ranked those two-attribute sets based on the results. We repeated this process of adding attribute one by one until acquiring the best set of attributes. In addition, attributes having multi-categories were recreated as dummy variables for each category to study if any category of them has stronger impact on the model. The best set of attributes selected for the model includes;

- Counties
- Pavement Conditions: Unknown, Dry, Wet or Snow/Ice
- Number of Tractor-Trailers
- Number of Pick-Up Vans
- PM Peak Hour Indicator: 1 if occurred in 4 PM ~ 6:30 PM; 0 otherwise
- Night Indicator: 1 if occurred in 8 PM ~ 6 AM; 0 otherwise
- Number of Shoulder Blockage
- Number of drivers/occupants injured
- Number of drivers/occupants killed
- Lighting conditions: Daylight, Dawn/Dusk, Dark-Lights on or Dark-No lights
- Collision Type-Head On indicator
- Collision Type-Head On Left Turn indicator
- Road-795: 1 if an incident occurred on I-795; 0, otherwise

Figure 5 illustrates the overall model performance. Considering the limit sample size, the developed model has performed satisfactorily for incidents of from 120 to180 minutes, and for these between 180 and 240 minutes. The model also performs well for incidents with duration less than 60 or longer than 300 minutes, though the available incident samples for these categories are relatively small. On the other hand, the model cannot yield the expected level of performance for incidents with duration between 60 and 120 minutes.



*Note: Percentages represent the proportion of correctly estimated/predicted incidents based on the developed NBC model.

Figure 5. Overall Model Performance

OVERALL FINDINGS

The overall findings from the available Maryland data set are summarized below:

1. When developing RBTM, 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, due to their differences in nature and associated factors that may contribute to the variation in the resulting duration.
3. 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.
4. 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*.
5. Complex geometric features compounded with the location factor may also contribute to an increase in the incident duration. For instance, exits on I-495 and I-270 generally cause a longer duration than those the same type of incidents. This was reflected in several MNL models with negative coefficients of the related variables
6. The lighting condition is one of the significant factors contributing to the duration of fatality involved incidents, implying that such a factor may also be related to incident severity.
7. The duration of a fatality involved incidents is found to be correlated significantly with its collision type. Among various types of collisions, head on and head on left turn collisions were identified as the most critical factor.

CONCLUSIONS AND FUTURE RESEARCH

This paper has illustrated the set of models developed under the data constraints for estimating the incident duration. The proposed model is hybrid in nature which can be separated into: non-fatality and fatality-involved incidents. To estimate the duration of incidents in the former category, we have proposed the use of the Rule-Based Tree Model as a primary model, which is supplemented with a discrete model to capture more embedded relations. Due to the concern of sample size and extremely wide range of distribution, we have further explored the use of Naïve Bayesian Classifier for estimating the duration of fatality-involved incidents. Through the model development, numerous findings associated with relations between incident duration and its contributing factors were discovered.

With the proposed model, one can construct efficient advanced traveler information for motorists under non-recurrent traffic congestion. Furthermore, this projected incident duration will enable responsible traffic agencies to estimate the approximate range of delay and queue distance and thus inform the en-route drivers of traffic congestion in a timely manner with VMS, and assess if any detour operators or control actions are needed. Drivers with better informed traffic information when encountering an incident can then make a proper route choice decision with less anxiety, which may consequently increase their compliance to suggestions or guidance by responsible traffic agencies.

REFERENCES

(1) Transportation Research Board (TRB), *Special Report 209: Highway Capacity Manual*, National Research Council, Washington, D.C., 2000

(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, August 1987, pp. 375-395

(3) D. Nam, and F. Mannering, "An Exploratory Hazard-Based Analysis of Highway Incident Duration", *Transportation Research – A*, Vol. 34A, No. 2, 2000, pp. 85-102

(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, November/December 1997, pp. 459-466

(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, 1995, pp. 113-138

(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) Wei, C-H and Lee, Y, "Sequential forecast of incident duration using Artificial Neural Network models", *Accident Analysis & Prevention*, Elsevier, 2007, pp. 944-954

(9) Qi, Y and Teng, H, "An Information-Based Time Sequential Approach to Online Incident Duration Prediction", *Journal of Intelligent Transportation* Systems, Taylor and Francis Group, 2008, pp. 1-12

(10) T. Hill, and P. Lewicki, Statistics: Methods and Applications

(11) Khorashadi, P. E. "Analysis of Driver Injury Severity: Logit Models of Truck Involvement/Truck Causation." Ph.D. dissertation, 2003

(12) Koppelman, F. S. and Bhat, C. "A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models.",2006

(13) Ben-Akiva, M. and Lerman , *S.R. Discrete choice analysis: Theory and application to travel demand,* MIT Press, Cambridge, Massachusetts, 1985

(14) Washington, S., Karlaftis, M. and Mannering, F. *Statistical and Econometric Methods for Transportation Data Analysis*, CRC Press, Boca Raton, 2003