Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Policy implications from traffic accident analysis with a case study from Vietnam

An Minh Ngoc[a], Truong Thi My Thanh[b]

[a]*University of Transport and Communication, Lang Thuong, Dong Da, Hanoi 100000, Vietnam*
[b]*University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi 100000, Vietnam*

Abstract

An accident frequency study based on classification and regression tree analysis was performed to identify the most important factors influencing traffic accident. This approach made it possible to detect specific strategies and actions for reduction of serious accidents. The results of model indicate that among the predictors (vehicle type, cause, age, gender, location, time), vehicle type, cause and age have strong impact to the severity of traffic accident. Then the discussion about the policy implications is needed to identify and manage all available affected factors relating to traffic accident. Data are from Traffic Police Bureau in the period 2015-2017, located in Hanoi, Vietnam. Through the main results obtained, it can be concluded that the CART algorithm of the Decision Tree is a useful tool in identifying potential sites of accidents. Based on it, transport authorities could be able to develop the specific strategies and measures.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: Accident Frequency, Regression Model, Fatalities.

1. Introduction

According to the statistics of Vietnam, the impact of traffic accidents cost accounts for 2.5% of GDP (Hung, 2017). Thus, evaluating and analyzing the damages caused by traffic accidents has become more significant during the last decades. Although a number of policy makers, civil agencies, and researchers invest in a large amount of human resources, materials and financial supports every year to reduce traffic accidents, number of traffic accidents stay at a high level. It is necessary to find out the effective approaches to identify the influencing factors of traffic accidents.

So far, many regression accident-frequency models, for instance, linear regression models, logit and probit models or structural equation models have been widely used in the traffic safety area to forecast accident frequencies over specific factors, offering a significant improvement to safety programming approaches that rely exclusively on observed accident histories. However, most of regression models require the assumptions among the variables. If these assumptions are violated, or homoscedasticity of the residuals is violated, the erroneous analysis results would

2352-1465
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

be generated (Washington SP, 2003). When the requirements are not met, the analysis results may be biased, and remedial actions should be taken.

Classification and Regression Tree Analysis (CART), one of the most popular data mining techniques, was introduced by Breiman et al. (1998). It has been applied in many areas of business administration, medicine, industry, and engineering fields, etc. However, the application of CART model in analyzing traffic safety problems is limited. Therefore, the objective of this study is applying the CART technology to analyze the various factors influencing traffic accident frequency that occurred in Hanoi, Vietnam.

In the remainning of this paper, we explore the experimental context and data used for the analysis, Then, the methodology is presented. By using the collected data, CART is analyzed and evaluated. Finally, the results are obtained followed by policy implications and further research direction.

2. Experimental context

Hanoi is seriously facing transport problems such as traffic congestion, traffic accidents and air pollution. Currently, Hanoi covers an area of 334.5 km2 with a total population of over 7 million in 2017. Since 2011, about 7,739 traffic accidents have occurred in Hanoi, resulting in approximately 3,199 fatalities and 5,015 injuries. Traffic accidents that caused causalities varied from 1,027 in 2012 to 962 in 2017, sustaining a predominantly sideways trend. However, the statistics also showed an approximate 24% reduction effect in fatalities over this period, from 794 in 2011 to 603 in 2017. During that time, the number of automobiles increased from 4.8 million at the end of 2012 to 6.7 million in 2017—an increase of 28% in five years. This increase is forecasted to continue in the future, making traffic safety a major for concern and interest going forward.

With the raise of traffic accidents, national government and city governments have pursued a series of policies in the effort to improve traffic safety and to prevent accidental death and injury.

In Vietnam, traffic safety was first mentioned in 1995 at the Government Decree 49/1995/ND-CP, providing for administrative sanctions against violations of traffic safety. This decree was a significant milestone that marked the government's concern about traffic safety. Two years later, the National Traffic Safety Committee (NTSC) and local traffic safety committees in 63 provinces were established and traffic safety was recognized as one of the critical social issues in Vietnam. During the period of 1995-2014, regulations for penalties on traffic violations were continuously adjusted to increase fines for traffic violations, deter road users and prevent traffic accidents.

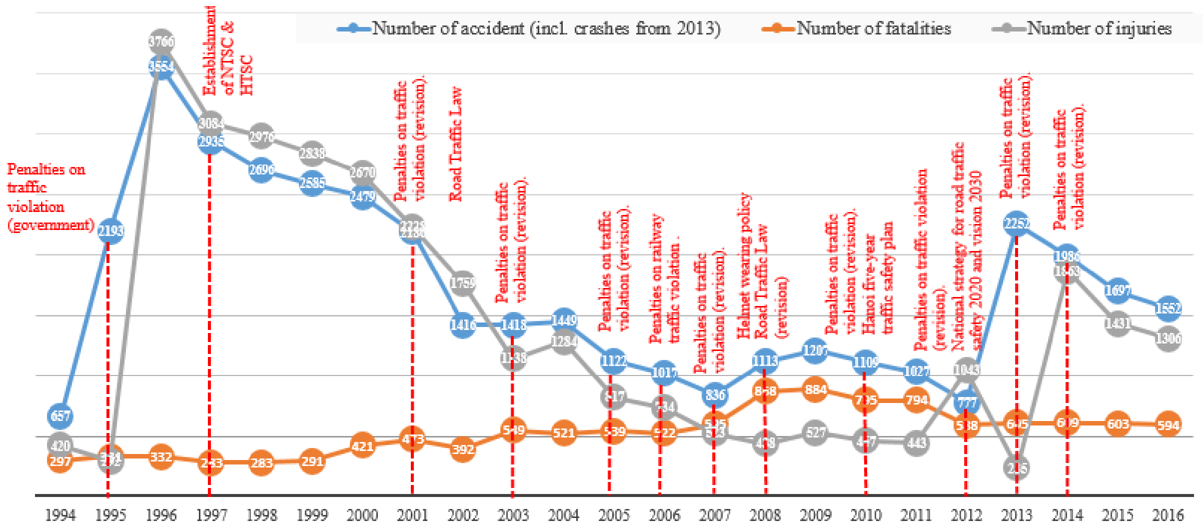

Fig. 1. Some milestones and traffic safety policies in Vietnam and Hanoi

At national level, helmet wearing is an important policy be applied that making significant contribution to the reduction of fatalities. This policy was promulgated in 2003 and officially applied in 2007, obligated motorcyclists to wear safety helmets when driving motorcycles on all types of roads.

At city level, the policy separating lanes for different types of vehicles was applied in 2009 in many roads in Hanoi. However, the implementation of this policy is failed since it conflicts with other road standards and having the protest from road users.

The accident data, specifically for this study, was collected in the duration of 2014 - 2017 from Hanoi Traffic Police Bureau. There were 1,054 accidents with six information fields, for instance, road location, time, vehicle involved in accidents, age and gender. The sample characteristics are represented in Table 1.

Table 1. Sample characteristics.

Characteristics	Statistics (N = 1,337)
Road segment	Section: n = 1185 (88.63%) Intersection: n = 152 (11.37%)
Time	6h00-17h59: n = 650 (48.62%) 18h00 – 5h59: n = 687(51.38%)
Vehicle	Motorcycle: n = 709 (53.1%) Commercial vehicle: n = 494 (36.9%) Passenger car: n = 97 (7.3%) Other: n = 37 (2.8%)
Age	Less than 18 years old: n = 19 (1.4%) 18-30 years old: n = 446 (33.36%) 30-45 years old: n = 589 (44.05%) Above 45: n = 283 (21.17%)
Gender	Male: n = 1139 (85.19%) Female: n = 198 (14.81%)
Cause	Violation: n = 1128 (84.29%) Non-violation: n = 209 (15.63%)

Traffic crashes were mostly recorded on road sections (88.63%) while on road intersections recorded only 11.37%. The distribution of crashes during night-time reached the highest number (51.38%). During 18pm to 22pm, male driver usually have party, meal with some wines or beers. Therefore, they often travel with less attention. This negative habit may increase the rate of traffic accident.

Motorcycle-related accidents was 53.1%, meanwhile car-related accidents accounted for 36.9%. Accidents relating to other vehicles took a small proportion.

The road users in the age group 30-45 had the highest involvement in traffic crash (44.05%). The age group 18-30 caused the second highest proportion of traffic accident (33.36%). The proportion of traffic accident reduced in the group of age over 45. The age group 18-30 is the most active duration in the human lives. Demand for driving of this age group is higher than other groups. However, they are not mature enough. Normally, the driving demand is reduced at a higher age. They are also more mature than the other age group, hence, the number of traffic accident decreased.

The traffic accident data specified by gender showed that the proportion of male-related accidents (85.19%) was much higher than female-related ones (14.81%). There are two major reasons for this result. Firstly, women drive more carefully and better obey the traffic rule than men, while men's confidence of his skill made him more careless in driving that easily lead to the traffic accident. Secondly, men usually drive when there are two people (men and women) in a vehicle according to the custom of Vietnam. This might be a reason for a higher proportion of male-related accident than female-related ones.

Traffic violation was also a significant reason for accidents occurring. More than 82% of road accidents resulted from either traveling too fast in urban areas or exceeding speed limits on rural roads.

3. Methodology

Decision tree is one of important data mining techniques for the classification and prediction of class variables. Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts to use the model to predict the classes of objects (Han and Kamber, 2006). When the value off the target variable is discrete, a classification tree is developed, whereas a regression tree is developed for the continuous target variable. The CART method can be used to develop either type of tree.

The development of a CART model generally consists of three steps. The first step is tree growing. The principle behind tree growing is to recursively partition the target variable to maximize "purity" in the two child nodes. Then, the data in each child node are more homogeneous that those in the upper parent node. In the tree-growing stage, splits are generated at each internal node of the tree. The Gini index will be used to measure the contribution of each split in maximizing homogeneity.

Gini Index for a given note t:

$$\text{GINI (t)} = 1 - \sum_t [p\left(\frac{j}{t}\right)]^2 \qquad \text{With } p\left(\frac{j}{c}\right) = \frac{p(j,t)}{p(c)} = \frac{\pi(j)N_j(t)}{N_j * \sum \; p(j,t)}$$

Where: p(j/t) is the relative frequency of class j at note t
j is the number of target class
π(j) is prior probability for class j
P(j/t) is conditional probability of a case being in class j provided that is in node m
$N_j(t)$ is the number of cases of class j of node m
N_j is the number of cases of class j in the roof node.

GI is one measure the degree of purity of the node, so when GI is equal to zero, the node is pure (all the cases in the node have the same class). When CART is development the aim is to achieve the maximum purity in the nodes, so the best split is the one that minimizes GI. Following this procedure, the maximal tree that overfits the data is created. To decrease its complexity, the tree is pruned using a cost-complexity measure that combines the precision criteria as opposed to complexity in the number of nodes and processing speed, searching for the tree that obtains the lowest value for this parameter

The second step is tree pruning. Pruning is the mechanism of producing a series of simple trees by removing the important nodes. During the whole process of pruning, the smaller trees are created gradually, forming into a pruned tree series.

The third step is to select the tree of an appropriate size from the pruned ones. When applied to analyze a new database, an over-sized number would produce higher misclassification. When the sample number is not big enough, all the data are usually expected to be used to establish the tree, and cross validation evaluation method can be used to provide an error rate assessment for each sample as well as to establish the tree.

The Chi-square tests were used to establish whether an observed frequency distribution differs from the theoretical distribution

More details about CART analysis and application can be found in Breiman et al. (1998), Chang and Chen (2005)

4. Results analysis and policy implications

The accident data is stored with three types including C0 - accident free, C1 - injury, and C2 - fatality. There were other accident types, however, they were not significantly associated with traffic operation.

The factors leading to the number of accident include location, time, type of vehicle, road user's characteristics, and the cause of accidents. For the tree growing, some indicators were grouped into smaller categories to achieve a sufficient representation. The grouping is represented in Table 2. Group 1 is road segment, including segment and intersection. Group 2 relates to time features of which daytime is considered as one variable. Variables in Group 3 are divided into two important categories (motorcycle, car & others). The variable "age" is reduced into two in Group 4. Group 5 is gender, and Group 6 describes traffic violation.

Table 2. Variable Description

Variable	Description	Average	Min.	Max.	Std.Err.
Group 1: Road segment					
Section	1: Accident occurred at section 0: Accident occurred at intersection	0.88	0	1	0.32
Group 2: Time Feature					
Daytime	1: Accident on daytime 0: Accident on night-time	0.49	0	1	0.5
Group 3: Vehicles involved in accidents					
Motorcycle	1: Motorcycle 0: Automobile & others	0.53	0	1	0.5
Group 4: Age					
Less than 45 year	1: Less than 45 years old 0: From 45 years old	0.72	0	1	0.45
Group 5: Gender					
Male	1: Male 0: Female	0.92	0	1	0.27
Group 6: Cause					
Violation	1: Traffic violation 0: Non violation	0.85	0	1	0.36

The first step of CART model is to determine the variables highly important to predict the dependent. This is achieved by using the importance index.

Using this index, five variables having the greatest influence on accident severity are detected (see Table 3). The most important variable causing fatality is vehicle type (57.64% importance). The violation variable has 52.39% importance in the model. The next important variable is age (42.08%). Gender variables are less important with percentage of 39 %.

Table 3. Importance of the variables

Accident frequency	Vehicle type	Cause	Age	Location	Time	Gender
100%	57.64%	52.39%	42.08%	41.3%	41.22%	39%

Base on the important level of variables, the classification tree is produced by CART method. The splitting procedure is explained as follows. The initial splitting of node 1 is based on the vehicle type. If the involved vehicle is motorcycle, CART puts it on the left side, forming terminal node 1, otherwise it puts it on the right side as node 2. For terminal node 2, CART predicts that 85.2 % accident free occurs. It means that the probability of fatality when accidents occurred is very small when using car. However, it is still possible to cause severity when people using motorcycles in node 1. Hence, the variable of splitting accident frequency selects the location. If accident occurs on road section, node 3 is produced on the left side, while node 4 is on the right side to indicate the accident at intersection. Because the time impacts to severity of accidents, node 3 and 4 are divided into node 7 and 8, node 9 and 10 respectively. Considering the influence of gender on traffic accidents, node 9 and 10 are divided step by step until the terminal nodes. Following this procedure, the whole tree of accident frequency prediction can be obtained. The prediction accuracy of data is 55.6%.

As illustrated in Fig. 1, the first variable to be used as predictor is vehicle type (node 1). A motorcycle using results severity is dead in accident with a probability of 61.1%. The node 1 is divided by location variable. Accidents that occur on sections are more likely to be fatality than those on intersections. A reason could be observed that the speed on road sections is much higher than intersection.

With specific location, most of serious accidents (84.1%) happen during the nighttime (6.00 pm – 6 am). Unsurprisingly, younger drivers are more likely to be involved in fatality at night-time when there is a less traffic on the road.

The age group less than 45causes not only more accidents but also higher fatality rate comparing to the elder age group. This result coincides with previous researches (Evans, 1991; Evans, 2000; Wiliams and Shabanova, 2003; Abdel-Aty et al, 1999). On the age group 18-45, males make up a considerably higher proportion of fatalities.

Around 90% of driver deaths are men. It shows that male drivers are more likely to be seriously injured or killed because of having higher speed collision.

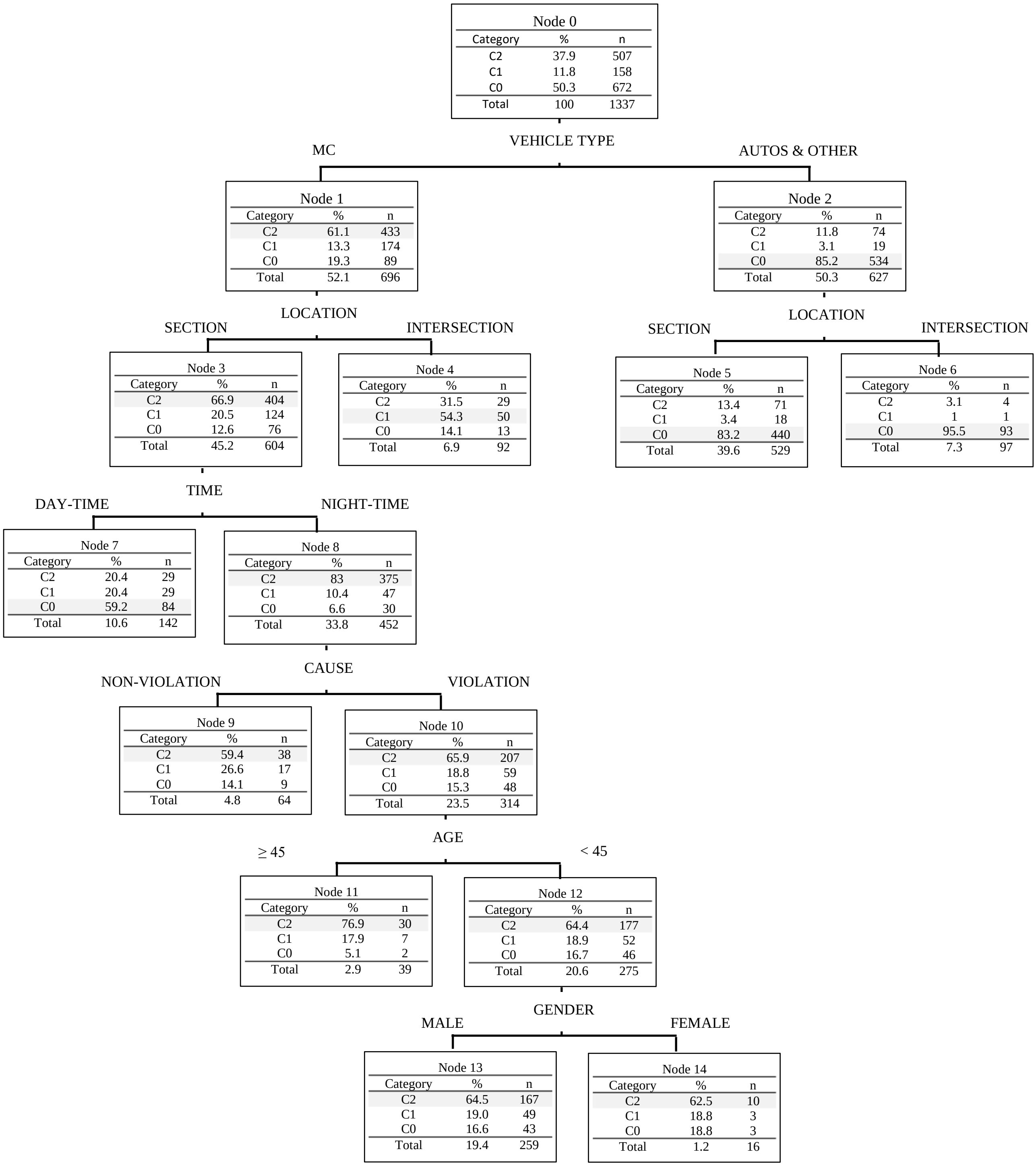

Fig. 1. Output of CART method

Following the paths created between root node and each terminal node at the models built, informative "if-then" rules are extracted. Interesting relationships of variables could be identified to explain the complication situations in traffic accidents. For example, specific traffic accidents face the following rules:

- NODE 2: IF (Vehicle using is AUTOMOBILE) THEN (accident frequency = ACCIDENT FREE).
- NODE 3: IF (Vehicle using is MOTORCYCLE and location is at SECTION) THEN (accident frequency = FATALITY
- NODE 4: IF (Vehicle using is MOTORCYCLE and location is at INTERSECTION) THEN (accident frequency = INJURE).
- NODE 7: IF (Vehicle using is MOTORCYCLE and location is at INTERSECTION and time is NIGHT) THEN (accident frequency = FATALITY).
- NODE 8: IF (Vehicle using is MOTORCYCLE and location is at INTERSECTION and time is DAY) THEN (accident frequency = ACCIDENT FREE).
- NODE 9: IF (Vehicle using is MOTORCYCLE Vehicle using is MOTORCYCLE and location is at INTERSECTION and time is at NIGHT and there is VIOLATION) THEN (accident frequency = FATALITY).

To reduce traffic accident and severity of accidents, there is a need to predict traffic accidents. With the above result analysis, the prediction process should be specified for Vietnam where motorcycles are dominant mode.

Table 4. Proposed strategies and detailed measures

Key Strategies	Description	Actions/Measures
Motorcycle Safety	Relief of human factor for motorcycle accidents	Reinforcement of aptitude test and safe driving education for motorcyclists in high risk group
		Enhancement of helmet quality
	Relief of vehicle factor for motorcycle accident	Life-cycle management of motorcycle
		Systemic management of motorcycle performance
Safe Road Infrastructure	ITS Based Traffic Safety Enhancement	Traffic violation notification system
		Improvement of automatic traffic violation enforcement system
	Road safety facility improvement	Road safety facilities installation
		Visibility improvement under poor condition (night)

- *Relief of human factor for motorcycle accidents*

It is the fact that the fatalities mostly happen in traffic accidents involved motorcycle. Therefore, searching the solutions for ensuring motorcycle safety should be placed on the first priority.

Solution for this situation might be built from frequently reviewing the driving ability and qualification of drivers to improve driver safety. It could also by introducing the driver's license system for drivers who are more likely to have traffic accident. For drivers who are in high-risk group, it is important to have a special inspection (frequent level). There are three components need to be inspected. The first component is special inspection items, including driving behavior, awareness of situation, judgment of risk, personality, eyesight, etc. The second component is qualification maintenance inspection items covering eyesight, attention, spatial judgment, visual memory, complex function, etc. The last component is safe driving education including identify the driving behavior of drivers through initial inspection, reinforce countermeasure capabilities in dangerous situations through experience education, carry out the corrective education such as individual face-to-face talk with a professional counselor.

Helmets are the most essential safety gear to protect the rider's head from fatality. However, 80% of the helmets used in Vietnam failed to meet the quality standard, particularly in impact absorption performance (according to 2012 WHO investigation).

- *Relief of vehicle factor for motorcycle accidents*

Motorcycle registration system in Vietnam is a once-off event. There is no feedback for existence history or scrapping. The following problems occur due to poor registration management of motorcycles.

- Possible to hold motorcycle without driving license, due to the lack of post registration management.
- Lack of basic data, which is necessary to establish the policy or goal of management for motorcycle.
- Difficult to impose the penalty based on license plate. There is mostly on-site police control.
- Quality management for motorcycle performance at government-level is impossible.

Motorcycle is the most important transport mode in Hanoi. If a systematic performance management procedure is introduced, the traffic safety level is significantly improved. Without legal performance inspection for motorcycle, motorcycle usage only depends on rider's discretion. Particularly, the management of the 'headlight performance' of the two-wheeled vehicle is important since the streetlight operation is poor in the suburban of Hanoi. However, a clear management is not performed. On the other hand, motorcycle overloading (more than two drivers are aboard) is a common situation in Hanoi. Hence, management of 'braking system performance' should also be important.

- *Traffic violation notification system*

Given the on-road behavior of people in Hanoi, they are failed to recognize the danger of traffic violation. Traffic accidents mostly occur because riders are careless in lane change or not carefully turning at intersection (based on traffic accident cause investigation).

Considering the data of traffic accidents, motorcycle accidents rate is quite high. It means that careless driving behavior of Hanoi citizen should be alarm accordance with the increase in vehicle number. Moreover, the accidents caused by over-speed and short of safe distance are growing, while considering the difference of speed due to various engines of different vehicles, it is essential to control and monitor these speeds. Currently, penalties are imposed as a sanction measure against traffic violations.

There is a lack of enforcement personnel and a strong opposition to the crackdown, which has not led to substantial improvements in traffic safety awareness. Therefore, it is necessary to strengthen the 'guidance' along with the crackdown. In other words, it is required to utilize road facility with high technology to detect traffic violation activities. Then, it alarms to violator and raise awareness about traffic violation activities to citizens. Through this approach, we can expect the changes of road user’s behavior for medium and long term.

Automatic enforcement system in Hanoi is also required because the face-to-face crackdown may cause the conflict with the violator or bribery. Also, it is difficult to implement due to the shortage of the manpower. To maximize the effect of automatic enforcement system, automatic detection of traffic violation is necessary. A high reliability system is required to prevent the conflict. When a signal violation detection system is operated in Hanoi, it will perform violation detecting and photographing together with a single camera. It is vulnerable to weather condition (torrential rain, fog) and the reliability may be reduced due to a wrong angle or low-definition image.

- *Road safety facility improvement*

It shows that 88.63% of accident occur in road segment. Among them, 45.2% of traffic accident caused by motorcycle users. It is necessary to install safety facilities on the roads to prevent traffic accidents and to eliminate defects in road environmental factors. Especially, large-sized vehicles and motorcycles are operated without any separated lanes.

To make safer roads in Hanoi, it is necessary to carry out various road safety improvement projects, for instance, frequent accident site improvement, dangerous road section improvement, safety facility maintenance projects. In addition, the safety management system is inadequate. A systematic analysis before project implementation, selection of business section and improvement alternative, a process of implementation and evaluation have not been implemented. Therefore, it is necessary to improve the road safety management procedures such as selection of project sections, diagnosis of causes, establishment of improvement alternatives, implementation and evaluation of effectiveness based on scientific and systematic risk assessment techniques.

Moreover, protective fence is needed to prevent the vehicle departing from the normal driving route during driving to enter to the opposite land or the sidewalk. Additionally, it is necessary to reduce the damage of the passengers and vehicles to minimum and returning the vehicle to the normal traveling direction. Also, it is important to attract the attention of a driver and prevent jaywalking.

The median barrier should composed by a separator and marginal strip. It prevents illegal movement from center line. It increases the traffic capacity, prevents misreading of the opposite lane, prevents glare caused by the headlights of the opposite vehicle at night, prevents U-turn and jaywalking. At elevated junction, the facilities securing the site of left-turn lane should be installed to protect the drivers having collision.

The motorcycle accident rate during 18:00 and 6:00 accounted for 51.16% of the total traffic accident at day-time and 33.8% at nigh-time by users. It shows that many vehicles driving at a high speed with poor visibility. Road safety facilities for improving visibility of lane marking at night are not sufficient enough. Therefore, delineation facilities need to be installed at the sides of the roads to specify the edge and the shape of the roads, guiding driver's sight during day and night time.

To prevent traffic accidents caused by low visibility of lane markings at night, the installation of delineation facilities is important in Hanoi. Traffic accidents are frequent and risks of accidents are high in this area.

Improve the safety of drivers by installing and repairing street lights at areas prone to night-time mortality accidents. Particularly, the lightings on national roads and local roads are poor. It requires immediate improvement.

5. Conclusions.

CART method allows classification based on crash severity and provides an alternative to parametric models because of their ability to identify patterns based on data, without the need to establish a functional relationship between variables.

Based on CART method, the key factors influencing traffic accidents were identified. Based on the findings, transport authorities could be able to develop specific strategies and measures.

The most important variables for the accident evaluation are vehicle type, cause, age, location, and time. Several interesting findings can be summed up. More than 50% of traffic accident has motorcycles involved. When there are motorcyclists involved in accident, the probability of fatality increases up to 61%. Accidents with the involvement of cars tend to have minor consequences, accident free. With a probability of 85.2%, this rule representing almost 47% of the total population. There is a higher probability of fatality in accidents that caused by violated male drivers of motorcycle in road segment are identified. The policy implication including enforcement of risky driving behaviours, installation of safety facilities, installation of traffic-monitoring cameras, introduction of a motorcycle safety system is recommended. Through multiple policy implications, road traffic crashes in Hanoi are expected to reduce in a relatively short time period.

Finally, CART method has several advantages inherent to non-parametric models. It does not require probabilistic knowledge of the study phenomena. There are no model assumptions or predefined underlying relationships between variables. Several advantages are found from tree models, for instance, the simplicity of interpreting the results for transport authorities, the graphic representation and the practicality of extracting "if-then" rules that can facilitate policy makers to choose appropriate strategies to focus on.

Acknowledgements

The authors are grateful to the Hanoi Police Bureau for providing the necessary data for this research. Without their support, it is impossible to complete this study.

References

Abdel-Aty, M.A., Chen, C.L., and Radwan, E., (1999). Using conditional probability to find driver age effect in crashes. Journal of Transportation Engineering, 125, 502-507.

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. London: Chapman & Hall/CRC; 1998.

Chang Ly, Chen WC (2005). Data mining of tree-based models to analyze freeway accident frequency. J Safety Res. 36(4): 365-375

Chang LY, Wang HW (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid Anal Prev; 38(5):10191027.

Evans, L., (1991). Older-drivers' risks to themselves and to other road users. Transportation Research Record: Journal of the Transportation Research Board, 1325, 34-41.

Evans, L., (2000). Risks older drivers face themselves and threats they pose to other road users. International Journal of Epidemiology, 29, 315-322.

Kashani, A. and Moheymany, A. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Safety Science 49, pp. 1314-1320

Lindley JA., (1987). Urban freeway congestion: quantification of the problem and effectiveness of potential solutions. ITE J. 57:27-32.

Hung, K.V (2017). Traffic safety strategies for Vietnam. Conference "Master Plan for Road Traffic Safety in Hanoi". 05.2017

Pakgohar A, Tabrizi RS, Khalili M, Esmaeili A.(2011). The role of human factors in incidence and severity of road crashes based on CART and LR regression: a data mining approach. Procedia Comput Sci. 3:764-769.

Washington SP, Karlaftis MG, Mannering Fl (2003). Statistical and econometric methods for transportation data analysis, NewYork: Chapman and Hall/CRC; 2003.

Williams, A.F., and Shabanova, V.I., 2003. Responsibility of drivers by age and gender, for motor-vehicle crash deaths. Journal of Safety Research, 34, 527-531.