Safety Science 50 (2012) 1034–1042

Contents lists available at SciVerse ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Effects of environment, vehicle and driver characteristics on risky driving behavior at work zones

Jinxian Weng, Qiang Meng *

Department of Civil & Environmental Engineering, National University of Singapore, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 27 July 2011
Received in revised form 7 November 2011
Accepted 4 December 2011
Available online 24 December 2011

Keywords:
Work zone
Risky driving behavior
Decision tree
Driver

ABSTRACT

This study aims to analyze the effects of environment, vehicle and driver characteristics on the risky driving behavior at work zones. A decision tree is developed using the classification and regression tree (CART) algorithm to graphically display the relationship between the risky driving behavior and its influencing factors. This approach could avoid the inherent problems occurred in the conventional logistic regression models and further improve the model prediction accuracy. Based on the Michigan M-94/I-94/I-94BL/I-94BR highway work zone driving behavior data, the decision tree comprising 33 leaf nodes is built. Bad weather, poor road and light conditions, partial/no access control, no traffic control devices, turning left/right and driving in an old vehicle are found to be associated with the risky driving behavior at work zones. The middle-aged drivers, who are going straight ahead in their vehicles with medium service time and equipped with an airbag system, are more likely to take risky behavior at lower work zone speed limits. Further, the middle-aged male drivers engage in risky driving behavior more frequently than the middle-aged female drivers. The number of lanes exhibits opposing effects on risky behavior under different traveling conditions. More specifically, the risky driving behavior is associated with the single-lane road under bad light or weather conditions while drivers are more likely to engage in risky behavior on the multi-lane road under good light conditions.

1. Introduction

Work zone is defined as a stretch of roadway for road maintenance or construction works. Work zones for road construction works usually last a long time (e.g., more than 3 days) while road maintenance works often occupy more than 1 h but less than 3 days (MUTCD, 2003). Speed limits must be posted in work zones and a part of lanes should be closed in order to guarantee the workers' safety. However, the presence of work zone could cause traffic congestion and create a more complex traffic environment for the traveling public. This is because traffic congestion could increase driver frustration, making drivers willing to engage in risky driving behavior in an effort to bypass delays (Maze et al., 2000). Although obstruction on a road may also increase the risky driving behavior, its features are quite different from those of work zones. For example, the obstruction may not last a long time because it may be quickly removed. Work zones pose unique challenges to the drivers' health and safety and this study concentrates on the analysis of risky driving behavior in work zones. Hereafter, the risky driving behavior is referred to as the following behavior: frequently speeding, aggressive lane changing, careless/negligent/reckless driving, driving too fast, failing to give way to pedestrians, disregarding traffic control signal and using an improper lane.

In spite of recent improvements on the work zone safety management, the high likelihood of severe work zone crashes is unacceptable. Risky driving behavior is still a major reason for the high likelihood of severe crashes at work zones. For example, a study conducted by Bai and Li (2006) reported that 92% of work zone crashes in Kansas are associated with the risky driving behavior. The current safety measures and policies cannot adequately reduce risky driving behavior (Hirsch, 2003; Mayhew and Simpson, 2002; Mayhew, 2007; Senserrick et al., 2009). Because of these reasons, there is a critical need to completely understand how the risky driving behavior is affected by the environment, vehicle and driver characteristics at work zones.

A number of studies have been conducted to identify the factors that can significantly affect the risky-taking behavior (Harre et al., 2000; Rhodes et al., 2005; Oltedal and Rundmo, 2006). Among these studies, the univariate statistical and multivariate regression methods are applied to compare the distributional difference between risky driving behavior and different groups (e.g., gender, age). However, it should be noted that the univariate statistical techniques only allow analysis of a single factor at a time. It may give rise to biased or incorrect results by isolating a single factor for analysis while treating others as fixed because some factors affect risky

* Corresponding author. Tel.: +65 6516 5494; fax: +65 6779 1635.
E-mail addresses: wjx206@gmail.com (J. Weng), ceemq@nus.edu.sg (Q. Meng).

0925-7535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2011.12.005

behavior interactively with others in reality. One discrepancy of the multivariate regression methods is that there still exists an inherent problem. One assumption for the multivariate regression method is that variables should be independent and an increase of value of one variable can be compensated for by proportionally decreasing or increasing the value of another variable to yield the same utility. Nonetheless, this assumption is often violated in the driver behavior analysis. For instance, drivers on the single-lane road are more likely to take risky driving behavior under bad traveling conditions while they are less likely to take this behavior under good traveling conditions. Therefore, the multivariate regression method may not accurately depict the relationship between the risky driving behavior and its influencing factors.

Decision tree is one of the popular data mining techniques, and it has been applied recently as an alternative approach to characterize drivers' behavior. One advantage of this technique is that it can graphically depict the relationship between risky driving behavior and its influencing factors. More importantly, decision trees could avoid the inherent problems occurred in the multivariate regression models and provide high prediction accuracy. Therefore, the focus of this study is to analyze the relationship between the risky driving behavior and environment/vehicle/driver characteristics at work zones using the decision tree approach. A classification and regression tree (CART) algorithm is employed to reproduce a decision tree.

1.1. Literature review

Many studies have been conducted on the analysis of risky driving behavior by using the univariate statistical methods. Elliott et al. (2006) assessed young drivers' gender differences in the associations between substance use/environmental influences and high-risk driving behavior. Their results showed that those women who misuse alcohol have fewer risky-driving incidents than men who abuse alcohol. Vassallo et al. (2008) examined the co-occurrence of risky driving with a range of externalizing and internalizing problems based on the χ^2-test results. They found the concurrent and longitudinal associations between risky driving and substance use (alcohol, cigarette and marijuana use, binge drinking).

However, the univariate statistical techniques only allow analysis of a single factor at a time. It may give rise to biased or incorrect results by isolating a single factor for analysis while treating others as fixed because the causes leading to the risky behavior are often complicated by presence of multiple factors. To deal with this problem, multivariate regression analysis techniques are applied for the analysis of risky driving behavior. Paschall (2003) investigated the relationship between college attendance and indicators of risk related driving (e.g., drinking and driving, seatbelt use) among young adults by using the logistic regression technique. The results showed that college students are more likely to drink and drive but also more likely to wear safety belts than non-students. Oltedal and Rundmo (2006) explored the effects of gender and personality traits including anxiety, excitement seeking, aggression and irritability. Personality traits and gender are found to explain 37.3% of the variance in risky driving behavior. Rhodes and Pivik (2011) examined the relationships among age, gender and risky driving by developing a regression model. They found that the teen and male drivers have more likelihood of risky driving than the adult and female drivers.

Although multivariate regression techniques are able to investigate the effects of multiple factors on the risky driving behavior, there exists a variable interaction problem. In order to eliminate the variable interaction problem, some researchers (e.g., Fernandes et al., 2007) abandoned the interacting variables with high degree of correlations based on the correlation analysis results. However, there is still one discrepancy for the above variable correlation analysis method because the interacting factors may have no interaction effects under certain conditions. In addition, the effects of environment, vehicle and driver characteristics (e.g., road condition, work zone speed limit, vehicle age) on the risky driving behavior at work zones have not been fully examined in the previous literature.

1.2. Objectives and contributions

The objective of this study is to reproduce a decision tree to investigate the risky driving behavior at work zones. Because of the simplicity and high accuracy, the CART (classification and retrogression tree) algorithm is employed to reproduce the decision tree. The effects of environment, vehicle and driver characteristics on the risky driving behavior are finally examined.

The contributions of this study are twofold. First, this study makes an initial attempt to analyze the effects of environment and vehicle characteristics on the risky driving behavior. This study could provide useful references for traffic engineers to propose effective measures/policies to reduce risky driving behavior at work zones. Second, the developed decision tree could avoid the variable interaction issue. The graphical display in the tree makes the relationship between risky driving behavior and its influencing factors easily understood.

2. Decision tree

A decision tree is a flow-chart-like tree structure where the root node is at the top and the leaf nodes are at the bottom. In a decision tree, the root node contains all records and the tree grows through the test of partitioning data at the nodes. The outgoing branches of a node correspond to all the possible outcomes of the test at the node. The leaves indicate the groups. To figure out which group of a record belongs to, we can start at the root node of the tree and trace a path down the tree according to the features of the record.

There are two types of decision trees. When the target variable is categorical, the decision tree is called a classification tree. As the target variable is continuous, the decision tree is called a regression tree. Because the outcome of the target variable driving behavior is dichotomous (i.e., unrisky and risky driving behavior), the reproduced decision tree is actually a classification tree in this study. Because of the simplicity, the CART (classification and regression tree) algorithm addressed by Breiman et al. (1984) is applied to reproduce a decision tree. The CART algorithm consists of two steps: tree growing and tree pruning, as shown in Fig. 1.

2.1. Tree growing

The principle behind growing a decision tree is to recursively partition the target variable so that the data in descendent nodes are always purer than the data in the parent node. When the training data enters the root node of a decision tree, a test is performed to search for all possible splits for all variables using a splitting criterion, which measures the quality of each possible split. In the CART, the Gini-index is the splitting criterion for growing a classification tree and the variance reduction is used as the splitting criterion for a regression tree. Since driving behavior is a binary target variable, the Gini-index splitting criterion is thus adopted in this study to select which variable and split scheme to be used to best split the node. The detailed splitting procedure is illustrated below:

Step1. For a given node t, the node impurity $i(t)$ is calculated according to the definition of Gini-index, shown as follows:

$$i(t) = 1 - \sum_{j} p_j^2 \tag{1}$$