공학석사 학위논문

Impact of Driver Behavior and Vehicle Type on Safety under Lane Change Situation

운전자 행동과 차량 종류가 차로 변경 시 안전에 미치는 영향 분석

2022년 8월

서울대학교 대학원

공과대학 건설환경공학부

신 용 우

Impact of Driver Behavior and Vehicle Type on Safety under Lane Change Situation

지도 교수 김 동 규

이 논문을 공학석사 학위논문으로 제출함

2022년 6월

서울대학교 대학원
공과대학 건설환경공학부

신 용 우

신용우의 공학석사 학위논문을 인준함
2022년 8월

위 원 장 고 승 영 (인)

부위원장 김 동 규 (인)

위 원 이 청 원 (인)

Abstract

Lane changes are critical contributors to road traffic safety on highways. Among the safety indexes aimed to evaluate the risks of these lane changes, the lane−change risk index (LCRI) is used to determine the potential collision probability of a lane−changing vehicle group in lane−change situations. This paper estimates the impact of driver behavior and vehicle type on the LCRI, using individual vehicle trajectory data. I defined a subject vehicle and its surrounding vehicles (i.e., lead, lag, front and rear vehicles) as a lane−changing vehicle group in a lane change situation. Each of their vehicle type (i.e., truck, bus, car, and motorcycle) and driver behavior (i.e., aggressive, ordinary, and timid) are categorized for regression analysis. Driver behavior is classified through time−space deviations between each vehicle's trajectories and expected trajectories from Newell's car−following model. In addition, to consider the heterogeneity among the lanes, this paper uses a linear mixed model, which reflects fixed and random effects. And the latent class analysis was used to classify the lane−changing vehicle group into a number of groups reflecting the characteristics of vehicle groups. Three unique findings of the present study are that (i) I quantified and analyzed the complex interaction between vehicle type and driver behavior within the lane−changing vehicle group in the situation of changing lanes, (ii) I found that the influence of the vehicle type and driver behavior in the lane−changing vehicle group had great heterogeneity depending on the lane, using the random parameter model, and (iii) when the lane−changing vehicle group was classified, most of the variables were observed to be statistically significant within two distinct classes. The findings of this study are expected to provide detailed lane−change strategies for autonomous vehicles as well as to evaluate the causative factors for lane−change risk.

Keyword : Driver Behavior, Vehicle Type, Lane−Changing Vehicle Group, Lane Change Risk Index, Latent Class Analysis

Student Number : 2020−27547

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Lane changes are critical contributors to road traffic safety on highways. Traffic crashes related to lane changes account for 4% to 10% of total crashes in the United States (Lee et al., 2003) and 13% of injury crashes in Germany (Mann et al., 2008). During lane changing, a driver interacts with neighbor vehicles in the current and the target lanes, considering driver position, speed, and acceleration. Predicting the movements of vehicles in the lane-changing vehicle group is important for a safe lane change. However, some drivers could behave unexpectedly due to driver behavior and/or vehicle type, which require complex decision-making and risk evaluation to avoid a traffic crash.

The risk of lane change situations has been evaluated by surrogate safety measures (SSM) quantifying the potential of crash risks. Time-to-collision (TTC) is a well-known SSM, indicating the time remaining to avoid an accident if a vehicle continues to drive in the same direction and speed (Hayward et al., 1972). More sophisticated SSMs have been proposed to consider the complexity of lane change situations. The crash propensity metric (CPM) has been proposed to estimate the probability of a simulated conflict considering the uncertainties of drivers and vehicles (Wang et al., 2014). These uncertain (i.e., unpredictable) behaviors of a vehicle in a lane-changing vehicle group have been reported to significantly increase the crash risk of the lane-changing vehicle group in lane change situations (Joo et al., 2021). The lane-change risk index (LCRI) proposed by Park et al. (2018) evaluates the collision probability of the lane-changing vehicle group by incorporating the exposure time and the expected severity level of potential crashes (Park et al., 2018). The CPM and LCRI can quantitatively evaluate the crash risk of the lane-changing vehicle group in lane change situations. Further, to better understand and predict the crash risk of lane changes, factors affecting the risk to the lane-changing vehicle group should be investigated.

Among the various factors affecting lane change, this study focuses on those characterizing the lane-changing vehicle group, including four surrounding vehicles (i.e., front and rear vehicles in the current lane and lead and lag vehicles in the target lane). Because the interactions of surrounding vehicles can be attributed to the speed, acceleration (Ma et al., 2021), and the SSM of each vehicle (Weng et al., 2018), I investigated two factors, driver behavior (i.e., aggressive, ordinary, and timid) and vehicle type (i.e., truck, bus, car, and motorcycle) to characterize the lane-changing vehicle group. The heavy vehicle (e.g., truck and bus) has different lane-change and car-following behavior from those of ordinary vehicles due to their differences in driving ability (Moridpour et al., 2009, 2008). Also, they have more significant influences on adjacent vehicles (Moridpour et al., 2012, 2010). Mixing different types of vehicles on intercity highways increases the LC ratio (Gu et al., 2018). This mixture also contributes to enticing effects in the fast lane, especially in the left-most lane. Accordingly, the traffic characteristics of the mixed types of vehicles must be adequately investigated in order to comprehend their relationship with lane-changing decisions (Moridpour et al., 2008). Also driver behavior is an important factor affecting road traffic safety. Driver behavior such as aggressiveness and reaction pattern significantly influence car-following behavior and interactions within the lane-changing vehicle group (Chen et al., 2014, 2012). Aggressive driving behavior can increase roadway crash potential as well as reduce travel speed (Park et al., 2019). Especially, aggressive driving behavior is well known to be closely related to the severity and occurrence of collisions (Stephens et al., 2014).

Despite these impacts of aggressive driver behaviors and heavy vehicles on traffic situations, their impact on the crash risk of lane change has not been fully investigated. Park et al. (2019), for example, evaluated the crash potential index (CPI) in various aggressive driving events, based on a driving simulator and microscopic traffic simulation (Park et al., 2019). Their findings revealed that aggressive driving deteriorates safety performance represented by CPI. Although this study reported the significant impact of driver aggressiveness on lane−changing risk, the risk was not evaluated at the vehicle group level.

As a remedy, this study aims to evaluate the impact of driver behavior and vehicle type on the crash risk of the lane−changing vehicle group in the lane change situation. A car−following model to consider the driver's behavior has been proposed in several studies. Tan et al. (2017) proposed 'an aggressive car−following model' based on the Krauss model (Krauss et al., 1997) that defines a safe distance between vehicles (Tan et al., 2017). The aggressive car−following model considers the driving style using the situation when the safe distance breaks. Tang et al. (2014) suggested the full velocity difference (FVD) model that considers the driver's attributes such as vehicle's speed, acceleration, driver behavior (i.e., aggressive, neutral, and conservative) (Tang et al., 2014), which was classified through vehicle speed and acceleration. They also showed that the aggressive driver had much greater vehicle speed and acceleration than other drivers (i.e., neutral, conservative). Lastly, Laval and Lecelercq (2010) proposed the L−L model to consider driver aggressiveness in car−following model (Laval et al., 2010). Among the above models, I chose the L−L model for three reasons: (a) the L−L model is the simplest model based on Newell's model and considers the driving behavior using a single parameter and thus, it can minimize the effects of other parameters; (b) the L−L model could reproduce the traffic oscillation caused by the situation when lane−change frequently occurred (Zheng et al., 2011); (c) the NGSIM data used in this study was also utilized to verify the performance of the L−L model (Chen et al., 2012). Based on empirical vehicle trajectory data, I measured the aggressiveness of each vehicle's driver in the congested traffic (Chen et al., 2014, 2012).

Then, four surrounding vehicles and subject vehicles were characterized by driver behavior and vehicle type. A linear mixed model (LMM) was applied to estimate the effect of driver behavior and vehicle type on the crash risk measured by LCRI. Based on the estimation results of LMM, I found that (a) vehicle type of the subject and surrounding vehicles are significantly associated with the LCRI of a lane−changing vehicle group; (b) aggressive drivers in the front and lag vehicle tend to increase the LCRI; (c) for the effect of driver behavior and vehicle type, there exist significant random effects across the lanes.

Additionally, in this study, the lane−changing vehicle group was classified based on the attributes of vehicle groups. A method of classifying groups can be largely divided into a hierarchical method and a non−hierarchical method. The K−means method is a representative non−hierarchical clustering method, and it divides the clusters by grouping them based on the distance within the group. However, when analyzing categorical variables, there is a disadvantage in that it is difficult to calculate and analyze the distance between variables (Singh et al., 2011). Therefore, to solve these problems, this study intends to apply the Latent Class Analysis (LCA) based on the probabilistic model. A LCA is an analysis method that can explain differences between classes by stratifying groups with similar characteristics based on their probability of belonging to each class (Wu et al., 2021). Furthermore, it classifies the latent classes with similarities based on the observed variables and finds out the types of the class (Collins, L. M. & Lanza, S. T., 2010). As a result, a LCA showed the results of defining group−specific characteristics and classifying groups according to a gaussian distribution. Based on the estimation results of LMM classified by LCA, I found that (a) the lane−changing vehicle group was classified by the model according to group attributes such as average speed within the vehicle group and inter−vehicle gap. (b) most of the variables were statistically significant within two distinct classes; latent class 1 (38.6% of the data sample) and latent class 2 (61.4% of the data sample). (c) compared with the LMM classified by LCA and the existing LMM, it was confirmed that there was a difference between the internal groups.

These two model approaches were adopted for the following reasons. While LMM was used to reflect the effect across the lanes through random effect, the LCA was used to identify differences between internal groups.

The rest of this paper is organized as follows. First, I described the study site and methods for measuring driver behavior and LCRI in detail. Then, I discussed the findings from the estimation results of the existing LMM and the LMM classified by LCA. Last, I made concluding remarks and note plan for future research.

Chapter 2. Methodology

2.1. Data

Data description

This paper used US−101 data among NGSIM dataset (NGSIM, 2008). The NGSIM datasets are vehicle trajectory data extracted from videos. The sampling frequency for NGSIM trajectories is 0.1 seconds; each sample contains lateral and longitudinal positions, instantaneous speed, acceleration, vehicle length, and vehicle type (He, 2017). The US−101 dataset was collected from a section of the US−101 Highway of Los Angeles, California. This section is about 640 m long and contains six lanes. The data collection period is 45 min (i.e., from 7:50 am to 8:35 am on June 15, 2005).

In **Figure 2.1**, a lane−changing vehicle group in the lane change situation is defined as a total of four surrounding vehicles, including the front (V_F) and rear (V_R) vehicles in the current lane and the lead (V_{Le}) and lag (V_{La}) vehicles in the target lane, which affect risk when the subject vehicle changes lanes (Park et al., 2018; Ma et al., 2021; Li et al., 2020; Chen et al., 2021; Chen et al., 2019). Vehicle position/speed, vehicle type, and driver behavior of each vehicle in the lane−changing vehicle group are measured by individual vehicle trajectory data.

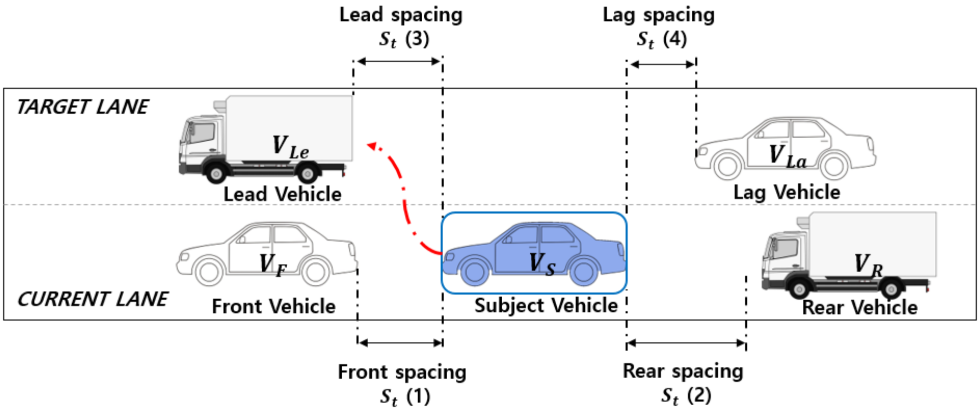

Figure 2.1 Definition of lane−changing vehicle group under lane change used in this study

Figure 2.2 illustrates the definition of lane change in this paper. Based on this definition, this study measures the starting point and ending point of lane change as well as lane−change time and distance (Park et al., 2018).

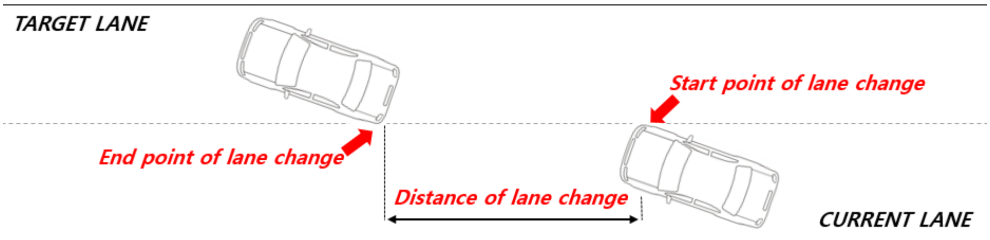

Figure 2.2 Definition of lane change used in this study

Table 2.1 shows a description of the dependent and independent variables selected for the regression analysis. In the models, the LCRI was used as a dependent variable indicating the lane change risk, and the remaining variables were used as independent variables. Traffic density was used as an indicator for traffic conditions; it was calculated by dividing the total number of vehicles observed when changing lanes by the length of the road (640 m). In NGSIM data, vehicles were classified as cars, trucks, buses, and motorcycles. However, the proportion of motorcycles among all observed vehicles is very small at 0.7%, and the motorcycle has a large difference in interaction with other vehicles during the lane change process. Therefore among them, I classified the vehicle types as automobiles and the trucks/buses that are heavy vehicles. Regarding driving behavior, the indicator of aggressiveness $\boldsymbol{\eta_i(t)}$ in Equation 8 is a continuous variable by definition. I had attempted to analyze the $\eta_i(t)$ as a continuous variable, or with categories of aggressive, timid, and ordinary, but both of them did not provide statistically significant parameter estimates for the aggressiveness. Therefore, I categorized the driver's aggressiveness into only aggressive and others (timid, ordinary). I only set the variable if a driver is aggressive. To represent the lane-changing vehicle group's safety condition during the lane change, I included the minimum TTC, lane change time, and the lane-changing vehicle group's average speed as independent variables.

TABLE 2.1 Description of the Independent and Dependent Variable

Variable Name	Explanation	Data Type	Ratio by Variable
LCRI (Dependent)	Lane change risk index	Numeric	
Traffic state			
Density	Indicator by dividing the total number of vehicles during lane change by 640 m	Numeric	
Driver's behavior			
Subject's behavior	1=Aggressive; 0=Otherwise (Timid, Ordinary)	Dummy	33.7%
Front's behavior	1=Aggressive; 0=Otherwise (Timid, Ordinary)	Dummy	35.7%
Rear's behavior	1=Aggressive; 0=Otherwise (Timid, Ordinary)	Dummy	33.7%
Lead's behavior	1=Aggressive; 0=Otherwise (Timid, Ordinary)	Dummy	37.21%
Lag's behavior	1=Aggressive; 0=Otherwise (Timid, Ordinary)	Dummy	36.77%
Vehicle's type			
Subject's type	1=Heavy vehicles (Bus, Truck); 0=Otherwise (Car, Motorcycle)	Dummy	1.47%
Front's type	1=Heavy vehicles (Bus, Truck); 0=Otherwise (Car, Motorcycle)	Dummy	2.75%
Rear's type	1=Heavy vehicles (Bus, Truck); 0=Otherwise (Car, Motorcycle)	Dummy	2.06%
Lead's type	1=Heavy vehicles (Bus, Truck); 0=Otherwise (Car, Motorcycle)	Dummy	2.56%
Lag's type	1=Heavy vehicles (Bus, Truck); 0=Otherwise (Car, Motorcycle)	Dummy	5.72%
Lane-changing situation			
Minimum TTC	Minimum TTC among the lane-changing vehicle group	Numeric	
Average speed	Average speed of the lane-changing vehicle group	Numeric	
Lane-change time	Time taken by subject vehicle to change lanes	Numeric	

2.2. Lane-Changing Risk Index (LCRI)

Vehicles driving along the road continue to interact with surrounding vehicles. In this paper, I defined these interacting vehicles as a lane-changing vehicle group in a lane change situation. Therefore, I measured the risk of the lane-changing vehicle group in lane-changing using the LCRI (Park et al., 2018), which estimates the collision risk, taking into account the exposure time and the expected severity level of potential crashes during lane-change. Individual vehicle trajectory data were used to calculate the LCRI.

Here, the concept of LCRI is provided based on the process of Park et al. (2018). The LCRI is based on the fault tree analysis (FTA), a widely used method for analyzing complex events caused by several reasoning factors (i.e., risk-severity level (RSL) and risk-exposure level (REL)). The FTA aims to identify the relationship between whole system failure and the failure of each system components (i.e., event failure); both RSL and REL are measured by the stopping distance index (SDI), which is an indicator determining the risk of rear collision on stopping sight distances (SSDs). In Figure 1, the front spacing $\boldsymbol{S_t(1)}$ in the current time step (t) is derived based on front vehicle and subject vehicle. SSD and SDI can be respectively obtained using Equation 1 and Equation 2. In Equation 1, V is the vehicle speed (kph), $\mathbf{f}$ is the coefficient of friction, $\mathbf{g}$ is the grade (decimal), and $\boldsymbol{t_r}$ is the perception-reaction time (2.5 seconds). The $\mathbf{f}$ and $\mathbf{g}$ are set as 0.28 and 1, respectively, based on the highway capacity manual (Highway Capacity Manual, 2000) and the average grade of the study site (Tan et al., 2017). In Equation 2, $\boldsymbol{SDI_t}(1)$ calculated by letting front vehicle $(\boldsymbol{V_F})$ be leading vehicle and subject vehicle $(\boldsymbol{V_S})$ be the following vehicle at time step t, $\boldsymbol{S_t}(1)$ is the gap between subject vehicle $(\boldsymbol{V_S})$ and front vehicle $(\boldsymbol{V_F})$ at time step t, $\boldsymbol{SSD_t^{V_S}}$ is stopping sight distance for subject vehicle $(\boldsymbol{V_S})$, $\boldsymbol{SSD_t^{V_F}}$ is stopping sight distance for front vehicle $(\boldsymbol{V_F})$ and $\boldsymbol{l_{V_F}}$ is length of front vehicle $(\boldsymbol{V_F})$:

$$SSD = \frac{V^2}{254 \times (f \pm g)} + t_r \times V \times 0.278 \tag{1}$$

$$\begin{cases} SDI_t(1) = f(x) = S_t(1) + SSD_t^{V_F} - SSD_t^{V_S} - l_{V_F} \\ SDI_t(1) > 0 \Rightarrow safe(0) \\ SDI_t(1) \le 0 \Rightarrow safe(1) \end{cases} \tag{2}$$

An SDI greater than zero indicates a situation in which $\boldsymbol{V_S}$ can safely stop when $\boldsymbol{V_F}$ suddenly stops. On the other hand, if SDI is less than zero, it is a dangerous situation, where $\boldsymbol{V_S}$ cannot properly evade a collision with the vehicle in front of it. Using the previously obtained SDI, I derived risk exposure level (REL) and risk severity level (RSL).

The degree to which a subject vehicle is exposed to a collision risk situation during a lane change is defined as an indicator of the increase in the likelihood of a collision, such as REL (Equation 3). Unsafe lane change duration (ULCD) is calculated by adding time steps with SDI less than zero and total lane change duration (TLCD) is calculated as total lane change time. Therefore, REL(i) is the proportion of unsafe lane−change duration among total lane−change duration. The risk severity level (RSL), developed to reflect situations in which the severity of the collision increases due to a relatively high collision rate, increases the absolute value of SDI if the subject vehicle's speed is faster than the leading vehicle's speed. Therefore, the absolute value of such SDI can be defined as the potential crash severity. In Equation 4, $\boldsymbol{SDI_{MAX}^{obs}}$ is the observed maximum SDI during TLCD; $\boldsymbol{SDI_{cri}}$ is theoretical maximum SDI:

$$REL = \frac{ULCD}{TLCD} \tag{3}$$

$$RSL = \frac{SDI_{MAX}^{obs}}{SDI_{cri}} \quad (4)$$

$\boldsymbol{SDI_{cri}}$ was obtained assuming that the speed of the following vehicle was the fastest among all the vehicles in the data and the interval between leading and following vehicles was 0 m. Therefore, RSL(i) is defined by observed maximum stopping distance index (SDI) divided by theoretical maximum SDI.

This study assumed that the REL(i) and RSL(i) represent the components of the probability of event failure in the perspective of duration and severity, respectively. This is the key assumption that the LCRI can be viewed as a probability interpretation. Based on this assumption, the system failure (i.e., probability of not performing a safe lane change) can be defined by integrating all event failures regarding adjacent vehicles (**φ(i)**). An event failure (**φ(i)**) between the subject vehicle and the i−th surrounding vehicle (V_F, V_R, V_{Le}, V_{La}) could be obtained by the product of REL(i) and RSL(i) as in **Eq. (5)**.

$$\varphi(i) = \mathrm{REL}(i) \times \mathrm{RSL}(i) \quad (5)$$

The system failure (i.e., probability of not performing a safe lane change) can be defined by integrating all event failures regarding adjacent vehicles (φ(i)). Since the system failure occurs even if only one event failure occurs, the system failure can be calculated as **Eq. (6)**. The system failure is the fault of the top−level event, and event failure ($\varphi(i)$) is the cause event that spreads up system failure. Event failure is defined as the failure to maintain safe interactions between the subject vehicle and the adjacent vehicles, which causes the probability of not performing a safe lane change.

$$\mathrm{LCRI} = \varphi(V_S) = 1 - \prod_{i=1}^{4}[1 - \varphi(i)] \quad (6)$$

2.3. Driver Behavior Measurement

Figure 2.3 shows the trajectories of vehicles in the time−space diagram. At this time, the traffic volume indicates how many vehicles have passed at a fixed location. In addition, the time difference between the leading and the following vehicles at the current location is defined as time headway, calculated as the reciprocal of the traffic volume.

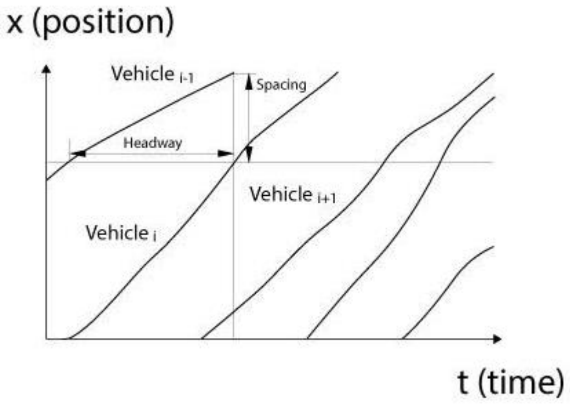

Figure 2.3 Time−space diagram of vehicle trajectory

In **Figure 2.4**, the Greenshield model showing the relationship between traffic volume and density is called the fundamental diagram. At this time, the model that simplified this model into line segments is the basic diagram of Newell's triangle model. And where the time heady (τ) is the reciprocal of traffic volume.

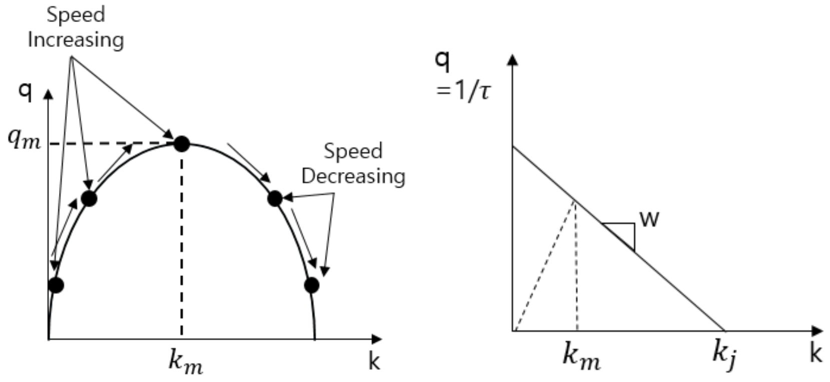

Figure 2.4. Greenshield model (Traffic Volume – Density) and Newell Triangle model (Traffic Volume – Density)

This study extracts driver behavior from individual vehicle trajectory data based on the definition of Chen et al. (2012, 2014). Specifically, I applied Newell's car−following model (Newell, 2002) and triangular foundation, to obtain $\boldsymbol{\tau_i(t)}$ in Figure 2.5, which is the wave trip time between two continuous congested vehicles. Then, the driving behavior is classified by $\boldsymbol{\eta_i(t)}$ (Equation 8) representing the reaction patterns to the shockwaves or traffic oscillation:

$$\tau = \frac{1}{kw} \tag{7}$$

where **w** is the wave speed and **k** is the jam density.

$$\eta_i(t) = \frac{\tau_i(t)}{\tau} \tag{8}$$

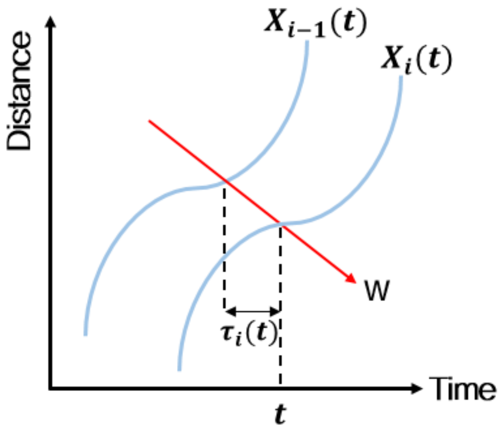

Figure 2.5. Measurement of $\tau_i(t)$

The τ_i(t) in **Equation 8** is the actual wave trip time, and τ is the average of $\tau_i(t)$ given by **Equation 7** and **Figure 2.5**. Therefore, for τ, I took the average of $\tau_i(t)$ across all drivers in the sample. Before a driver experiences a stop−and−go disturbance, $\eta_i(t)$ is essentially consistent across time, but deviates as the driver accelerates and decelerates, depending on the driver's reaction pattern to lead vehicle. Thus $\eta_i(t)$ describes the time−dependent behavior of the driver. Here, among the samples, I observed that driver behavior follows a consistent pattern across traffic oscillations. This allows it to determine whether the driver is timid or aggressive based on the trajectories of Newell's car−following model.

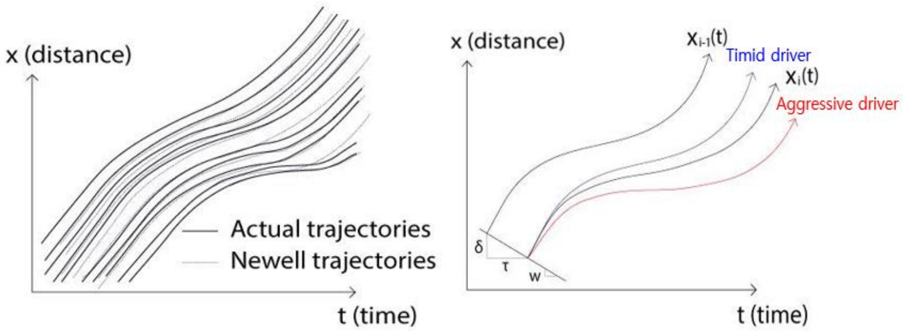

Figure 2.6. Examples of deviation from Newell trajectories and Driver behavior

In **Figure 2.6**, if the follower is a timid driver, a driver reacts early when the leader decelerates or late when the leader accelerates ($\boldsymbol{\eta_i \geq 1.1}$). At the time, the time and distance difference between the leader and the follower increases. In other situations, if the follower is an aggressive driver, a driver reacts late when the leader decelerates or early when the leader accelerates ($\boldsymbol{\eta_i \leq 0.9}$). The time and distance between the leader and the follower are reduced.

2.4. Linear Mixed Model (LMM)

The NGSIM data show a significant difference in the ratio of heavy vehicles along lanes (Oh et al., 2015). Different heavy vehicles–to–lane ratios on these roads can affect road safety due to reduced capacity and behavior characteristics (e.g., lower acceleration performance and wider spacing than ordinary vehicles). In addition, various lane–specific factors, such as desire speed, proximity to on–/off–ramps and traffic composition, can affect lane change situations (Duret et al., 2012).

To consider the effects of these lane–specific properties, in this study, I used a linear mixed model (LMM) that included random–effect parameters in the linear regression model (LRM) (Kim et al., 2020; Bates et al., 2014). The LMM is a statistical model that describes a continuous dependent variable as a combination of fixed effect and random effect (Pinheiro et al., 2000; Snijders et al., 2011; Gelman et al., 2006) as in Equation 9:

$$\begin{aligned} &\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon}, \\ &\boldsymbol{\gamma} \sim N(0, \boldsymbol{G}), \\ &\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{R}). \end{aligned} \tag{9}$$

where $\boldsymbol{Y}$ is the dependent variable, X is the design matrix for fixed–effects, Z is the design matrix for random– effects, $\boldsymbol{\varepsilon}$ is a residual error and follows a distribution with mean 0 and variance R, $\boldsymbol{\beta}$ is a vector of fixed–effect parameters, and $\boldsymbol{\gamma}$ is a vector of random–effect parameters and follows a distribution with mean 0 and variance G. To compare the LMM with LRM, I fitted the model using maximum likelihood rather than restricted maximum likelihood that is widely used for LMM. More details about the estimation procedure for LMM are described in (Bates et al., 2014).

Here, I considered the LMM with random slopes, assuming that some parameters have random effects according to lanes since the effect of the vehicle type and driver behavior on LCRIs could vary by lane. Random effects of intercept are excluded since it was estimated to be insignificant. All the cases of lane changes are grouped separately (e.g., from lane 1 to lane 2, from lane 2 to lane 1, and from lane 2 to lane 3). Since there are six lanes, the total cases of lane groups are 10.

2.5. Latent Class Analysis (LCA)

In general, cluster analysis is a statistical method of grouping individuals with similarities. In this case, it is very important to search for a natural cluster by the similarity of individuals or variables in multivariate data. Also, such cluster analysis can be divided into a hierarchical method and a non-hierarchical method. The K-means method is a representative non-hierarchical clustering method, and it divides the clusters by grouping them based on the distance within the group. However, when analyzing categorical variables, there is a disadvantage in that it is difficult to calculate and analyze the distance between variables (Singh et al., 2011). Meanwhile, existing cluster analyzes may not find optimal clusters because the analyst subjectively determines the number of clusters. To solve this problem, latent class analysis, which uses variables to classify the entire group into subdivided groups, can be used. The latent class analysis proposed by Lazarsfeld (1950) and Lazarsfeld & Henry (1968) has a model involving observed variables and latent variables (Lazarsfeld, 1950; Lazarsfeld & Henry et al., 1968). The categories of latent variables are called latent classes. The latent class analysis, which is a case of model-based clustering for multivariate data, assumes that the observation value of each individual is extracted from one of several groups, and models each group with a probability distribution. In this analysis method, the data is modeled by assuming that the data of each individual is extracted from a mixture of a finite number of different distributions. At this time, it was assumed that the probability distribution for each group followed a Gaussian distribution.

In addition, the commonly used model selection method uses all variables, not just those useful in determining the number of groups. In this case, in LCA, only useful variables are selected and the data are modeled to determine the appropriate number of groups. In this study where many variables exist, LCA was used to classify the groups for the as above reasons.

The algorithm of latent class analysis (Kim, 2013) can be expressed as **Equation 10**.

$$P(Y_{i1} = y_{i1,\dots,} Y_{in} = y_{in}) \tag{10}$$
$$= \sum_{r=1}^{c} P(L_i = r) P(Y_{i1} = y_{i1,\dots,} Y_{in} = y_{in} | L_i = r)$$

Equation 10, $P(L_i = r)$ represents the probability that the individual i belongs to class r. For example, suppose I have two classes.

$$Y = (y_1, y_2) = Y_1 = (y_{11}, y_{21}), \dots, Y_n = (y_{1n}, y_{2n}) \tag{11}$$

$$P(r_{u,t}) = \pi_1 P(r_{u,t} | L_1) + \pi_1 P(r_{u,t} | L_2) \tag{12}$$

Equation 11, it is assumed that y1 and y2 are independent. This is the process of finding the probability that y1 and y2 are 1. The probability of belonging to which group is obtained.

Here, in order to select the latent variable for LCA, I determined the group characteristics in the lane-changing vehicle group defined through a review of previous studies. In a lane change situation, through previous studies that considered vehicle groups rather than a single vehicle, the characteristics of vehicle groups could be determined as average speed, gap between vehicles, and lane change time(Chen et al., 2021; Maiti et al., 2017). Among these characteristics, the attributes that classify the vehicle group in consideration of the number of groups and interpretability were determined by the average speed and the inter-vehicle gap.

Chapter 3. Results

3.1. Linear regression model and Linear mixed model

In this paper, I applied two different modeling approaches: linear regression model (LRM) and linear mixed model (LMM). At this time, the model was applied by considering only the dangerous situation, excluding the sample when the lane change risk is 0. The LRM statistically provides a functional relationship between dependent and independent variables in the observed data to analyze their existing associations. I verified the validity of the LRM by testing whether the error term follows normal distribution using the Shapiro-Wilk normality test (Shapiro et al., 1965). The null hypothesis of the normality test is that the data are normally distributed, and the result for my data shows that the error term follows normal distribution with p-value of 0.2418 (i.e., it cannot reject the null hypothesis). The LRM cannot take into account intra-class correlation, leading to errors in which the standard error of parameter estimates is underestimated. To address this issue, I applied the LMM that explains data containing random effects as well as fixed effects.

The bottom of Table 3.1 provides the measure for model fit such as Akaike information criterion (AIC) and Log-likelihood to compare the two models. These measures show that the LMM outperforms the LRM at a 95% level of confidence based on a likelihood-ratio test. This result indicates that there are significant random effects among the lanes in measuring LCRI, and those effects can be represented by parameters for driver behavior and vehicle type. Therefore, I analyzed the estimation results focusing on LMM. I included all variables including the insignificant ones in the model to provide all information for my model specification.

3.1.1 Linear Mixed Model (Fixed Effect)

TABLE 3.1 Estimation Results of LRM and LMM

Dependent Variable : Lane Change Risk Index (LCRI)						
Fixed Effect	**Coefficients**		**Standard Errors**		***t*-Statistics**	
	LRM	**LMM**	**LRM**	**LMM**	**LRM**	**LMM**
Intercept	1.44**	1.39**	0.64	0.61	2.24	2.26
Traffic State						
Density	8.25***	7.81***	2.71	2.58	3.04	3.03
Driver Behavior						
Subject's behavior	-0.09	-0.05	0.19	0.18	-0.49	-0.28
Front's behavior	0.14	0.17*	0.18	0.29	0.76	2.60
Rear's behavior	0.17	0.19	0.19	0.18	0.90	1.08
Lead's behavior	-0.05	-0.11	0.19	0.18	-0.25	-0.63
Lag's behavior	0.26*	0.28*	0.18	0.20	2.46	2.41
Vehicle Type						
Subject's type	0.30	0.48**	0.72	0.69	1.41	2.69
Front's type	2.17***	2.34***	0.43	0.41	5.03	5.70
Rear's type	0.25	0.26*	0.54	0.56	0.46	2.47
Lead's type	0.87	0.71	0.61	0.57	1.43	1.23
Lag's type	-0.43	-0.35*	0.34	0.32	-1.30	-2.11
Lane-Changing Situation						
Minimum TTC	-10.34***	-9.76***	1.10	1.06	-9.39	-9.25
Average speed	-0.10***	-0.11***	0.03	0.03	-2.96	-3.42
Lane-change time	0.30***	0.32***	0.10	0.09	3.09	3.50
Random Effect	**Standard Deviation**		**Variance**			
Rear's type	0.42		0.18			
Front's behavior	0.68		0.46			
Lag's behavior	0.27		0.07			
	LRM	**LMM**	**Groups**		10	
Number of Observations	291	291				
AIC	1054.409	1050.912				
	Df	**Loglik**	**Df**	**Chisq**	**Pr(>Chisq)**	
LRM	16	-511.20				
LMM	19	-506.46	3	9.50	0.02**	

Notes: LRM is linear regression model; LMM is linear mixed model; Df is degree of freedom;
$^{*} = p < 0.1$; $^{**} = p < 0.05$; $^{***} = p < 0.01$

First, traffic density indicating traffic states shows significant positive effects on LCRI at the 1% significance level. In other words, the higher the density of vehicles in the lane, the greater the risk of the lane change. During lane change, higher density leads to a narrower space between the leading and following vehicles, and this spacing is critical for increasing LCRI. Also, the estimated coefficient is the highest among the parameters with significant positive effects. This result shows that traffic states are the dominant factor for the risk of collisions in the lane change, which comes from the increases in overall collision risk (Kononov et al., 2012, 2011). The significant impact of traffic states also has been reported in the analysis of actual lane−change collisions (Pande et al., 2006).

Regarding the lane−changing situations, lane−change time is significantly and positively associated with the LCRI, indicating that the longer the lane−change time of the vehicle in the lane, the greater the risk of the lane change. The average time taken to change lanes in this paper is 3.25 seconds, but the maximum is 12.8 seconds. This abnormal situation of a long−time lane change comes from the failure and retry of lane change, leading to drastic deceleration and close gap, which increases the risk of lane changes. According to Cao et al. (2013), lane change times are longer when maneuvers are more dangerous or when the interaction between the subject vehicle and surrounding vehicles is complex (Cao et al., 2013). In the higher traffic density situation, the lane change time was longer because it became difficult and dangerous to perform the lane change action (Toledo et al., 2007). Meanwhile, minimum TTC and average speed in the lane−changing vehicle group during a lane change show significant negative coefficients at the 1% significance level for LCRI. Considering the definition of TTC, a high TTC means a longer collision time with the leading vehicle, so the risk of lane change would be reduced at this time. At this time, the minimum TTC means the minimum value among the TTC values of vehicles belonging to the lane−changing vehicle group, not simply the TTC value of individual vehicles. Through this, it is possible to consider the effect of the deceleration of the leading vehicle on the overall lane−

changing vehicle group. Minimum TTC is also an indicator of criticality since it indicates a near−accident (Vogel et al., 2003). In the NGSIM data, most traffic is in congested or near−congested conditions. Therefore, a high average speed of lane−changing vehicle group indicates the relaxation of congestion, which reduces the lane−changing risk.

Vehicle types excluding lead's vehicle in the lane−changing vehicle group have shown significant effects for lane−change risk. Furthermore, except for the lag vehicle, the heavy vehicle in the lane−changing vehicle group tends to cause higher lane−change risk than a car and motorcycle at the 10% significance level. In particular, when the front vehicle is a heavy vehicle, the value of the coefficient is the highest, indicating the greatest impact of the front vehicle's type. The characteristics of a heavy vehicle, which takes up more road space and has a lower driving performance, such as a lower deceleration and braking capability than lighter vehicles, causes risky simulations to the surrounding vehicles, particularly in congested conditions (Al−Kaisy et al., 2005). In addition, the number of lane changes per lane occurs more frequently when more heavy vehicles are present in the lane, and the frequent lane changes can increase lane−change risk due to the unexpected other lane change (Moridpour et al., 2015). Therefore, the risk of lane change of the lane−changing vehicle group can increase if the heavy vehicles are included. On the other hand, when the lag vehicle is a heavy vehicle, it appears to reduce the risk of lane change. This may be due to the fact that the front space gap of heavy vehicle is wider in order to provide brakes safely than that of a car, and the speed of the heavy vehicle is slower than that of the surrounding vehicles (Moridpour et al., 2015; Zhao et al., 2018). The length of heavy vehicle and the difficulty for truck drivers to see vehicles in adjacent lanes also would be the causes for the risk of collision during lane−change (Khattak et al., 1998). Therefore, when a vehicle in the current lane changes lanes, the lane−changing risk could be reduced because the space between heavy vehicle and the front vehicle is maintained to be wider.

For driver behavior, the variables indicating significant and positive association were front's behavior and lag's behavior. This means that the risk of lane change increases if the driver's behavior of the front vehicle or lag vehicle is aggressive. Aggressive driving is a dangerous driving event that is likely to cause a traffic crash, which threatens other drivers via acts such as close driving and sudden deceleration and acceleration (Stephens et al., 2014). The aggressive follower reacts late when the leader slows down and early when the leader accelerates, thus reducing the time and distance between the leader and the follower. Therefore, the aggressive driving characteristics of the front vehicle (before lane change) and lag vehicle (after lane change), which pay the most attention to changing lanes, narrow the gap between space and time with the vehicle in front. Previous studies also reported that aggressive driving behavior could increase roadway crash potential (Park et al., 2019), and it is closely related to the severity and occurrence of collisions (Stephens et al., 2014).

3.1.2 Linear Mixed Model (Random Effect)

Some parameters for the vehicle type and driving behavior exhibit significant random effects across the 10 lane groups. Other insignificant random parameters with zero-variance were excluded in the model estimation. The coefficients of fixed effects are significantly changed by the LMM considering random effects. For example, the positive coefficient of the front's type increases 7.83% from 2.17 to 2.34 in the LMM compared with LRM, and the negative coefficient of minimum TTC increases 5.6% from −10.34 to −9.76. Also, the LMM reveals some significant parameters that were not significant in the LRM, including front's behavior, subject's type, rear's type, and lag's type; thus, richer insights for lane change risk can be provided by LMM. These changes come from the different estimation methods of LMM that estimates group-specific effects and LRM that ignores all of the information about the groups.

The variables with significant random effects across lanes are rear's type, front's behavior, and lag's behavior. The large standard deviation of random effect can be interpreted as a variable's effect heavily influenced by the lane. To represent the estimated random effect, I calculated conditional mode as shown in Table 3.2. The conditional mode represents a group-level value of the random effects, which is a conditional mean effect of the variable given groups (Bates et al., 2014). The coefficients for fixed effect and standard deviation of random effects indicate that the rear's type and the front's behavior can have significant proportions of both negative and positive values, and it is confirmed by the conditional mode as in Table 3.2. For the rear's type, the heavy vehicles tend to decrease the risk of lane change from 4, 5 to 5, 4, unlike to other lanes. Since the lane 4 and 5 are the driving lane for the heavy vehicles, other vehicles may maintain the spacing when they encounter the heavy vehicles. However, in the other lanes where the desired speed is higher than lanes 4 and 5, the other vehicle may narrow the spacing to overtake the heavy vehicles. For the front's behavior, the change from lane 1 to lane 2 has significantly greater positive effects than other lanes, while other lanes showed negative and positive effects without specific patterns. Since lane 1 is the fastest lane, the following subject vehicle may have a tendency to reduce spacing due to the high desired speed, and this impact could be amplified by the aggressive driving behavior of the front vehicle.

TABLE 3.2 Conditional Mode of Random Effects

Group	**Current Lane**	**Target Lane**	**Rear's Type**	**Front's Behavior**	**Lag's Behavior**
1	1	2	0.00	1.22	0.12
2	2	3	0.10	-0.11	-0.05
3	3	4	0.00	-0.72	-0.09
4	4	5	-0.12	-0.23	-0.11
5	5	6	0.00	0.48	0.04
6	2	1	0.00	-0.02	0.07
7	3	2	0.28	-0.31	-0.12
8	4	3	0.01	-0.39	-0.13
9	5	4	-0.14	0.45	0.32
10	6	5	0.02	-0.37	-0.04

The large standard deviation of random parameters and the changes in the coefficient and significance of fixed effects highlight the relevance of grouping variables. These results can be supported by prior studies that reported the significant random effects among the lanes to take into account the ratio of heavy vehicles (Oh et al., 2015) and the impact of on- and off-ramps per lane (Duret et al., 2012) on NGSIM data.

3.2. Latent Class Analysis (LCA)

In this paper, I compared the LMM classified into two classes by LCA and the existing LMM. The LMM is used to consider random effects among the lanes, whereas LCA is used to identify unobserved classes. When comparing the two models, the sign of the existing LMM parameters and the LMM parameters classified by LCA may be opposite. For example, in the existing LMM, the effect of variables is the result for the total data sample. But in the LMM classified by LCA, the effect of variables is the result for only a part of the data sample.

The bottom of Table 3.3 provides to determine the optimal number of classes in the latent class analysis, IC (Information Criteria), which are indices of fitness, are considered. Among the log−likelihood functions, the most commonly used methods are AIC and BIC. Therefore, in this paper, the number of latent class was determined in consideration of the AIC, the number of groups, and interpretability.

TABLE 3.3 Selection of the number of classes (avg_speed, inter−vehicle gap)

Class	Model fit		Classification rate(%)				
	AIC	BIC	1	2	3	4	5
2	3812.985	3882.778	0.386	0.614			
3	3835.215	3886.641	0.385	0.454	0.161		
4	3841.588	3889.747	0.425	0.108	0.296	0.171	
5	3929.281	3962.341	0.019	0.268	0.183	0.062	0.467

The smallest number in the AIC value was determined as the optimal cluster value. The lowest AIC is when the number of class is two. Total variables were observed to be statistically significant within two distinct classes; latent class 1 (38.6% of the data sample) and latent class 2 (61.4% of the data sample).

The **Figure 3.1** shows the range of the common latent factor used to classify the class respectively.

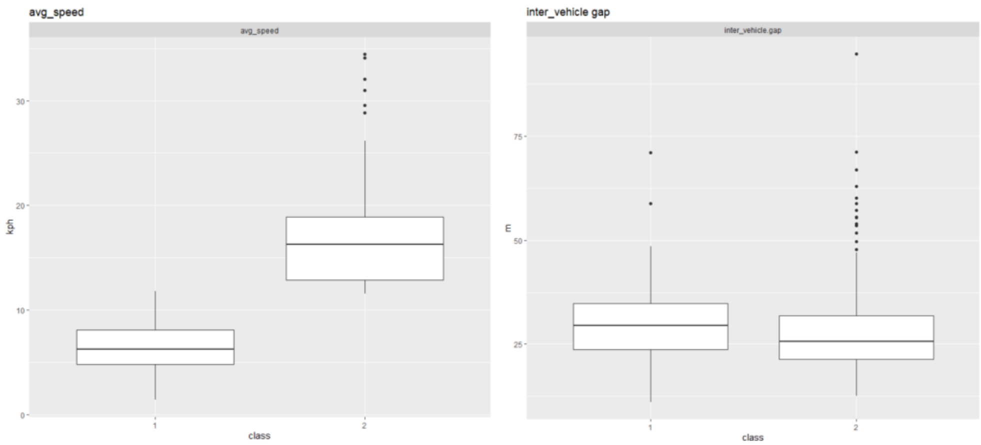

Figure 3.1 Box plots of average speed and inter−vehicle gap for two classes

Table 3.4 also shows the range of common latent factors for each specific class. The characteristics of the classes can be defined through the average of each value. In the case of class 1, it can be seen that the gap between the vehicle is wide, and the average speed in the vehicle group is low. On the other hand, in the case of class 2, it can be seen that the gap between the vehicle is narrow, the average speed in the vehicle group is high. Since the length of the section where the data was collected was 640 m, there was no significant difference in the gap between vehicles between the two classes.

TABLE 3.4 Average speed and inter−vehicle gap by classes (Range)

Class	**Class Probability**	**Gap distance (Mean) (m)**	**Average speed (Mean) (kph)**
1	0.386	11.14 ~ 71.01 (30.11)	1.40 ~ 11.78 (6.55)
2	0.614	12.50 ~ 94.69 (28.38)	11.54 ~ 34.42 (17.46)

In **Table 3.5**, as explained earlier, in the LMM classified into two classes by LCA, I focused on the variables with opposite signs for each class or value differences in coefficient values. In addition, since I tried to identify the differences between internal groups through the attributes of the vehicle group, I concentrated on the vehicle type and driver behavior variables within the vehicle group for the analysis.

First, among the driver behavior variables, lag's behavior showed opposite sign coefficients at the 5% significance level for LCRI by class. For lag's behavior in class 2, the aggressive driving tends to increase the risk of lane change, unlike in class 1. In the vehicle group with a high average speed and a close gap between vehicles, the higher speed of the lag vehicle results in a narrower space for the vehicle to change lanes. At this time, the risk when changing lanes may increase due to the urgency of the driver to complete the lane change (Chen et al., 2021).

Heavy vehicles except for the front and lead vehicle showed significant results for the risk of lane change at the 10% significance level in both classes. In addition, in both classes, when front and lead vehicle were excluded, heavy vehicles tend to cause a higher risk of lane change than cars and motorcycles. According to Zhao et al. (2018), when heavy vehicles maintain a narrow space from the vehicle in front, the lane-change risk is higher because heavy vehicles take longer to decelerate to avoid a collision compared to cars (Zhao et al., 2018).

The changes and the value differences in the coefficient highlight the differences between internal groups.

3.2.1 LMM classified by LCA

TABLE 3.5 Estimation Results of LMM classified into two classes by LCA

Dependent Variable : Lane Change Risk Index (LCRI)						
Fixed Effect	**Coefficients**		**Standard Errors**		***t*-Statistics**	
	Class1	**Class2**	**Class1**	**Class2**	**Class1**	**Class2**
Intercept	1.99**	2.45***	1.14	0.68	-1.75	3.61
Traffic State						
Density	5.96***	4.17**	4.40	2.60	4.99	2.10
Driver Behavior						
Subject's behavior	-0.44	0.05**	0.26	0.19	-1.68	-2.26
Front's behavior	0.05***	0.02**	0.37	0.23	2.69	2.07
Rear's behavior	0.63	0.18	0.27	0.19	1.35	0.99
Lead's behavior	-0.07	-0.07	0.27	0.19	0.27	0.39
Lag's behavior	-0.44**	0.25**	0.67	0.24	2.06	1.99
Vehicle Type						
Subject's type	0.99**	12.21**	0.82	2.15	2.22	-5.68
Front's type	1.31	2.55**	0.82	0.43	1.60	5.96
Rear's type	0.28**	0.37**	1.22	0.50	-2.04	2.74
Lead's type	0.80	0.35	0.99	0.59	0.81	0.59
Lag's type	0.09**	0.79**	0.43	0.35	-2.83	-2.19
Lane-Changing Situation						
Minimum TTC	-8.96	-8.78	1.78	1.11	-5.02	-7.93
Average speed	0.01	-0.26	0.03	0.04	0.41	-6.12
Lane-change time	-0.06	0.53	0.09	0.08	-0.66	6.78
Random Effect	**Standard Deviation**		**Variance**			
Rear's type	0.02	0.01	0.01	0.65		
Front's behavior	0.69	0.37	0.47	0.14		
Lag's behavior	1.71	0.41	2.93	0.17		
	Class1	**Class2**	**Groups**		10	
Latent Class Probability	0.386	0.614				
Loglik	**-125.34**					

Notes: LCA is Latent class analysis; LMM is linear mixed model; Df is degree of freedom;

$^{*} = p < 0.1$; $^{**} = p < 0.05$; $^{***} = p < 0.01$

Chapter 4. Conclusions

Lane changes are important for road traffic safety on highways. Thus far, the risk of lane−change situations has been assessed by the surrogate safety measures (SSM), which quantifies the probability of collision risk, and various safety indicators exist in this regard. Among these indicators, this study adopts LCRIs (Park et al., 2018) to evaluate the potential collision probability of the lane−changing vehicle group by integrating the exposure time and the expected severity level of potential crashes in lane−change situations. I investigated the impact of driver behavior and vehicle type on the LCRI using vehicle trajectory data collected in congested conditions. For a lane−changing vehicle group consisting of four adjacent vehicles and a subject vehicle, I characterized the lane−changing vehicle group using vehicle type (i.e., heavy vehicle or car/motorcycle) and driver's behavior (i.e., aggressive or ordinary/timid) of each vehicle in the lane−changing vehicle group. I employed an LMM to identify fixed and random effects of driver behavior and vehicle type and found that heavy vehicles and aggressive driver behaviors had different effects on crash risks of lane−changing vehicle group, depending on the role in lane change situations. Also the LCA was used to identify differences between internal groups. The characteristics of each class were defined by LCA, and the results of comparing the existing LMM and LMM classified by LCA showed that some heavy vehicles and aggressive driver behaviors had different effects on the collision risk of the lane−changing vehicle group by class.

The findings of this study provide interesting insights into the lane−change risk of a lane−changing vehicle group. I quantified and analyzed the complex interaction between vehicle type and driver behavior within the lane−changing vehicle group in the situation of changing lanes through LCRI. Additionally, using the random parameter model LMM, I found that the influence of the vehicle type and driver behavior in the lane−changing vehicle group had great heterogeneity depending on the lane. Based on the results of the LMM, the risk of lane changes was significantly associated with the vehicle type and driver behavior which is consistent with the expectation in the literature (Park et al., 2019; Stephens et al 2014; Al-Kaisy et al 2005; Moridpour et al., 2015; Zhao et al., 2018). The results show that when the vehicle type is a heavy vehicle, there is a significant correlation with LCRI. It implies that the lane change and car-following behavior of a heavy vehicle are different from an ordinary vehicle according to the difference in driving ability increase the possibility of a collision in a lane change situation. In particular, when the front vehicle is a heavy vehicle, the coefficient value is the highest, indicating that the type of the front vehicle has a greater influence than other adjacent vehicles. Meanwhile, the aggressive driver behavior had a significant positive coefficient for LCRI only for front and lag vehicles. The random-effects estimated by LMM were significant for rear's type, front's behavior, and lag's behavior among the groups of 10 lanes.

And when comparing the LMM classified into two classes by LCA and the existing LMM, I focused on the variables with opposite signs for each class or value differences in coefficient values. In addition, since I tried to identify the differences between internal groups through the attributes of the vehicle group, I concentrated on the vehicle type and driver behavior variables within the vehicle group for analysis. For lag' behavior in class 2, the aggressive driving tends to increase the risk of lane change, unlike in class 1. In the vehicle group with a high average speed and a close gap between vehicles, the higher speed of the lag vehicle narrows the space between the vehicles, which increases the risk when the driver changes lanes. In addition, in both classes, when front and lead vehicle were excluded, heavy vehicles tend to cause a higher risk of lane change than car and motorcycle. Due to the low driving performance of heavy vehicles, maintaining a narrow space with the vehicle in front increases the risk.

The proposed models evaluate the lane-change risk of the lane-changing vehicle group using LCRI. This measure can be applied to evaluating network safety performance for lane change considering the attributes of vehicle group, the proportion of heavy vehicles, and aggressive drivers in a roadway and help to develop countermeasures such as reducing aggressive driving. Also, my model evaluates the risk at the level of the vehicle group rather than the single-vehicle. In the connected vehicle (CV) technology, the lane-changing vehicle group to improve the safety of cruise control has been defined similar to my study (Ma et al., 2020). Therefore, my findings would be particularly valuable in the CV environment. Lane change in mixed traffic with human-driving and autonomous vehicles (AVs) is a difficult decision for AVs due to complex interactions with human-driving vehicles. This would be particularly valuable in the mixed traffic situations with human-driving and autonomous vehicles (AVs). Lane change in mixed traffic is a difficult decision for AVs due to complex interactions with human-driving vehicles. My findings from empirical data are valuable not only for understanding the current driving situations but also for designing the lane-change strategy of autonomous vehicles that should consider the complex interactions with human-driven vehicles with different behavior.

To further advance my findings, several limitations should be considered in future research. First, although I conducted a comprehensive analysis for 45 min of vehicle trajectory data, various traffic conditions, environmental conditions, and geometric conditions were not discussed because other trajectory data were not available. Recently collected vehicle trajectories from the advancement of traffic surveillance technologies such as unmanned aerial vehicles (UAV) (Kim et al., 2019) will enable further studies for various conditions. Through observation in a wider section, it may be possible to reflect the effects of more distant surrounding vehicles. Also, methodology such as Kaplan-Meier estimation was used to effectively estimate censored data related to missing data (Ozguven et al., 2008). Using this method can increase the number of samples and broaden the scope of the analysis. Second, this study measures the risk of lane-changing using LCRI (Park et al., 2018) and driver behavior based on Newell's car−following theory (Chen et al., 2012). Other measures for lane−changing risk and driver behavior can be applied to my model specification for generalizing my findings. Especially for driver behavior, driving simulation (Park et al., 2019) and micro traffic simulation (Habtemichael et al., 2014) would be the promising methods for more detailed consideration of behaviors. Last, the risk evaluation of a lane−changing vehicle group would be an important subject for studying mixed−traffic conditions. The driving data collected from autonomous vehicles can be analyzed to evaluate the risk of a lane−changing vehicle group in the lane change. Further, comparing the risk with the human−driving vehicle could provide valuable insights into designing lane−change strategies.

Bibliography

Lee, S. E., E. C. B. Olsen, and W. W. Wierwille., 2003. A Comprehensive Examination of Naturalistic Lane Changes. Report DTNH 22-00-C07007, Task Order 4. NHTSA, U.S. Department of Transportation

Mann, M. Benutzerorientierte Entwicklung und fahrergerechte Auslegung eines Querführungsassistenten., 2008. Cuvillier Verlag

Hayward, J. C. Near-Miss Determination Through Use of a Scale of Danger. In Highway Research Record 384, HRB, National Research Council, Washington, D.C., 1972. 24–34.

Wang, C., and Stamatiadis, N. Derivation of a New Surrogate Measure of Crash Severity. Transportation Research Record: Journal of the Transportation Research Board, 2014. 2432:37–45.

Joo, Y.-J., H.-C. Park, S.-Y. Kho, and D.-K. Kim. Reliability-Based Assessment of Potential Risk for Lane-Changing Maneuvers. Transportation Research Record: Journal of the Transportation Research Board, 2021. 03611981211010800.

Park, H., C. Oh, J. Moon, and S. Kim. Development of a Lane Change Risk Index Using Vehicle Trajectory Data. Accident Analysis and Prevention, 2018. 110:1–8.

Ma, Y., Z. Lv, P. Zhang, and C. Y. Chan. Impact of Lane Changing on Adjacent Vehicles Considering Multi-Vehicle Interaction in Mixed Traffic Flow: A Velocity Estimating Model. Physica A: Statistical Mechanics and its Applications, 2021. 556:125577.

Weng, J., Du, G., Li, D., and Yu, Y. Time-Varying Mixed Logit Model for Vehicle Merging Behavior in Work Zone Merging Areas. Accident Analysis and Prevention, 2018. 117:328–339.

Moridpour, S., M. Sarvi, G. Rose, and E. Ramsay. Variables Influencing Lane Changing Behaviour of Heavy Vehicles. In Proc. 31st Australasian Transport Research Forum, 2008. 1–15.

Moridpour, S., G. Rose, and M. Sarvi. Modelling the heavy vehicle drivers' lane changing decision under heavy traffic conditions. Road & Transport Research: A Journal of Australian and New Zealand Research and Practice, 2009. 18:49–57.

Moridpour, S., G. Rose, and M. Sarvi. Effect of Surrounding Traffic Characteristics on Lane Changing Behavior. Journal of Transportation Engineering, 2010. 136:973–985.

Moridpour, S., M. Sarvi, G. Rose, and E. Mazloumi. Lane-Changing

Decision Model for Heavy Vehicle Drivers. Journal of Intelligent Transportation Systems, 2012. 16:24–35.

Guo, M., Wu, Z., and Zhu, H. Empirical Study of Lane-Changing Behavior on Three Chinese Freeways. PloS one, 2018. 13:e0191466.

Chen, D., J. Laval, Z. Zheng, and S. Ahn. A Behavioral Car-Following Model That Captures Traffic Oscillations. Transportation Research Part B: methodological, 2012. 46:744–761.

Chen, D., S. Ahn, J. Laval, and Z. Zheng. On the Periodicity of Traffic Oscillations and Capacity Drop: The Role of Driver Characteristics. Transportation Research Part B: methodological, 2014. 59: 117–136.

Park, S., C. Oh, Y. Kim, S. Choi, and S. Park. Understanding Impacts of Aggressive Driving on Freeway Safety and Mobility: A Multi-Agent Driving Simulation Approach. Transportation Research Part F: Traffic Psychology and Behaviour, 2019. 64:377–387.

Stephens, A. N., and M. J. M. Sullman. Development of a Short Form of the Driving Anger Expression Inventory. Accident Analysis and Prevention, 2014. 72:169–176.

Tan, F., Wei, D., Zhu, J., Xu, D., and Yin, K. An aggressive car-following model in the view of driving style. Canadian Journal of Civil Engineering, 2017. 44.10:775-782.

Krauss, S., Wagner, P., and Gawron, C. Metastable states in a microscopic model of traffic flow. Physical Review E, 1997. 55.5:5597.

Tang, T. Q., He, J., Yang, S. C., and Shang, H. Y. A car-following model accounting for the driver's attribution. Physica A: Statistical Mechanics and its Applications, 2014. 413:583-591.

Laval, J. A., and Leclercq, L. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2010. 368.1928:4519-4541.

Zheng, Z., Ahn, S., Chen, D., and Laval, J. Freeway traffic oscillations: microscopic analysis of formations and propagations using wavelet transform. Procedia-Social and Behavioral Sciences, 2011. 17:702-716.

Singh, K., Malik, D., & Sharma, N. Evolving limitations in K-means algorithm in data mining and their removal. International Journal of Computational Engineering & Management, 2011, 12(1), 105-109.

Wu, Y., Hu, H., Cai, J., Chen, R., Zuo, X., Cheng, H., and Yan, D. Applying latent class analysis to risk stratification of incident diabetes among Chinese adults. Diabetes Research and Clinical Practice, 2021. 174: 108742.

Collins, L. M. & Lanza, S. T., 2010, Latent class and latent transition analysis: With applications in the social, behavioral and health sciences.

US Department of Transportation. NGSIM – Next Generation Simulation. 2008. http://www.ngsim.fhwa.dot.gov/. Accessed October 31, 2008.

He, Z. Research Based on High-Fidelity NGSIM Vehicle Trajectory Datasets: A Review. Research Gate, 2017. 1–33.

Li, M., Z. Li, C. Xu, and T. Liu. Short-Term Prediction of Safety and Operation Impacts of Lane Changes in Oscillations with Empirical Vehicle Trajectories. Accident Analysis and Prevention, 2020. 135: 105345.

Chen, Q., Huang, H., Li, Y., Lee, J., Long, K., Gu, R., & Zhai, X. Modeling accident risks in different lane-changing behavioral patterns. Analytic methods in accident research, 2021, 30: 100159.

Maiti, Santa, Stephan Winter, and Lars Kulik. "A conceptualization of vehicle platoons and platoon operations." Transportation Research Part C: Emerging Technologies 80, 2017, 1-19.

Chen, T., X. Shi, and Y. D. Wong. Key Feature Selection and Risk Prediction for Lane-Changing Behaviors Based on Vehicles' Trajectory Data. Accident Analysis and Prevention, 2019. 129: 156–169.

Highway Capacity Manual (HCM). Highway Capacity Manual 2000, Transportation Research Board, National Research Council, Washington, D.C., 2000.

Newell, G. F. A Simplified Car-Following Theory: A Lower Order Model. Transportation Research Part B: Methodological, 2002. 36:195–205.

Oh, S., and H. Yeo. Impact of Stop-and-Go Waves and Lane Changes on Discharge Rate in Recovery Flow. Transportation Research Part B: Methodological, 2015. 77:88–102.

Duret, A., S. Ahn, and C. Buisson. Lane Flow Distribution on a Three-Lane Freeway: General Features and the Effects of Traffic Controls. Transportation Research Part C: Emerging Technologies, 2012. 24:157–167.

Kim, E. J., D. K. Kim, S. Y. Kho, and K. Chung. Spatiotemporal Filtering Method for Detecting Kinematic Waves in a Connected Environment. PloS one, 2020. 15:e0244329.

Bates, D., Mächler, M., Bolker, B., and Walker, S. Fitting Linear Mixed−Effects Models Using Lme4. arXiv preprint arXiv:1406.5823, 2014.
Pinheiro, J. C., and Bates, D. M. Linear mixed−effects models: basic concepts and examples. Mixed−effects models in S and S−Plus, 2000. 3–56.
Snijders, T.A., and Bosker, R. J. Multilevel analysis: An introduction to basic and advanced multilevel modeling. sage, 2011.
Gelman, A., and Hill, J. Data analysis using regression and multilevel/hierarchical models. Cambridge university press, 2006.
Lazarsfeld, P. F. The logical and mathematical foundation of latent structure analysis. Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction, 1950. 192−412.
LAZARSFELD, P. F., AND HENRY, N. W. Latent Structure Analysis. Boston: Houghton Mifflin. 1968
Kim, Sora. “Variable Selection In Latent Class Analysis.” Master's Thesis., Sungkyunkwan University, 2013
Shapiro, S. S., and Wilk, M. B. An analysis of variance test for normality (complete samples). Biometrika, 1965. 52.3/4:591−611.
Kononov, J., Lyon, C., and Allery, B. K. Relation of flow, speed, and density of urban freeways to functional form of a safety performance function. Transportation research record, 2011. 2236:11–19.
Kononov, J., Durso, C., Reeves, D., and Allery, B. K. Relationship between traffic density, speed, and safety and its implications for setting variable speed limits on freeways. Transportation research record, 2012. 2280: 1–9.
Pande, A., and Abdel−Aty, M. Assessment of freeway traffic parameters leading to lane−change related collisions. Accident Analysis & Prevention, 2006. 38.5:936−948.
Cao, Xiaoying, William Young, and Majid Sarvi. Exploring duration of lane change execution. Australasian Transport Research Forum. 2013.
Toledo, T., & Zohar, D. Modeling duration of lane changes. Transportation Research Record, 2007. 71−78.
Vogel, K. A comparison of headway and time to collision as safety indicators. Accident analysis & prevention, 2003. 35.3:427−433.
Al−Kaisy, A., Younghan, Jung, and Rakha, H. Developing passenger car equivalency factors for heavy vehicles during congestion. Journal of transportation engineering, 2005.

131:514–523.
Moridpour, S., Mazloumi, E., and Mesbah, M. Impact of heavy vehicles on surrounding traffic characteristics. Journal of advanced transportation, 2015. 49:535–552.
Zhao, P., and Lee, C. Assessing rear-end collision risk of cars and heavy vehicles on freeways using a surrogate safety measure. Accident Analysis & Prevention, 2018. 113:149–158.
Khattak, A. J., Kantor, P., and Council, F. M. Role of adverse weather in key crash types on limited-access: roadways implications for advanced weather systems. Transportation research record, 1998. 1621.1:10-19.
Ma, Z., Huo, Q., Yang, X., and Zhao, X. Safety cruise control of connected vehicles using radar and vehicle-to-vehicle communication. IEEE Systems Journal, 2020. 14.3:4602-4613.
Kim, E.-J., H.-C. Park, S.-W. Ham, S.-Y. Kho, and D.-K. Kim. Extracting Vehicle Trajectories Using Unmanned Aerial Vehicles in Congested Traffic Conditions. Journal of Advanced Transportation, 2019.
Ozguven, E. E., & Ozbay, K. Nonparametric Bayesian estimation of freeway capacity distribution from censored observations. Transportation Research Record, 2008. 2061.1: 20-29
Habtemichael, F. G., and de Picado Santos, L. Crash risk evaluation of aggressive driving on motorways: Microscopic traffic simulation approach. Transportation Research Part F: Traffic Psychology and Behaviour, 2014. 23:101–112

Abstract

고속도로에서 주행을 하면서 운전자가 빈번하게 수행하는 차로 변경은 도로 교통안전과 교통 흐름에 큰 영향을 미치는 행위라고 할 수 있다. 차로 변경 시 위험을 평가하기 위한 안전 지표들 중에서 Lane Change Risk Index(LCRI)는 차로 변경 상황에서 차로를 변경하는 차량 그룹(lane-changing vehicle group)의 잠재적 충돌 가능성을 결정하기 위해서 사용했다. 본 연구의 목표는 개별 차량 궤적 데이터를 사용하여 LCRI에 대한 운전자 행동 및 차량 종류가 미치는 영향을 추정하는 것이다. 이를 위하여 차로 변경을 실시하는 차량과 그 주변 차량(차로변경 전 선행, 후행 차량과, 차로변경 후 선행, 후행 차량)들을 차로를 변경하는 차량 그룹으로 정의했다. 각각의 차량 종류(트럭, 버스, 자동차 및 오토바이)와 운전자 행동(공격적인, 보통 및 소심한) 변수들은 회귀 분석을 위해 분류되었다. 이 때 운전자 행동은 각 차량의 궤적과 Newell의 차량추종모델의 예상 궤적 간의 시공간 편차를 통해 분류했다. 또한 본 연구는 차로 그룹 간의 이질성을 고려하기 위해 고정 효과와 임의 효과를 반영할 수 있는 선형 혼합 모델을 사용하였다. 그리고 잠재 계층 분석법을 이용하여 차로를 변경하는 차량 그룹을 차량 그룹의 특성을 반영하여 여러 그룹으로 분류하였다.

본 연구의 결과는 다음과 같다. 먼저 차로 변경 상황에서 차로를 변경하는 차량 그룹 내의 차량 종류와 운전자 행동 사이의 복잡한 상호 작용을 정량화하고 분석했다. 또한 Random parameter model을 사용하여 차로를 변경하는 차량 그룹에서 차량 종류와 운전자 행동의 영향이 차로에 따라 큰 이질성을 가지는 것을 발견했다. 끝으로 차로를 변경하는 차량 그룹들을 분류했을 때, 대부분의 변수들은 두 개의 별개 집단 내에서 통계적으로 유의한 것을 관찰했다. 이러한 발견들은 자율주행차의 세부적인 차로 변경 전략을 제시하고 차로 변경 시 위험의 원인 요인을 평가하는 데 상당한 기여를 할 수 있음을 시사한다.

주요어 : 운전자 행동, 차량 종류, 차로를 변경하는 차량 그룹, 잠재 계층 분석
학번 : 2020-27547