

디지털 운행기록에 근거한 시내버스 운전자의 사고발생 예측모형 개발

Development for City Bus Driver's Accident Occurrence Prediction Model Based on Digital Tachometer Records

김정열* (Jung-yeul Kim) (Myongji University)
금기정** (Ki-jung Kum) (Myongji University)

요 약

본 연구는 시내버스 운전자의 실제 운행기록 정보를 토대로 사고발생 가능성을 내포한 운전자를 판단할 수 있는 모형개발을 목적으로 하였다. 본 연구를 위하여 사고발생 운전자 및 사고 미발생 운전자의 실제 운행기록 정보에서 교통사고와 관련된 유의변수를 도출하는 한편, 판별분석(Discriminant Analysis) 및 로지스틱회귀분석(Logistic Regression Analysis)을 적용하여 개발된 분류모형에 대한 모형간 정확도를 비교하였다. 또한, 개발된 모형을 다른 운전자들의 운행기록자료에 적용하여 모형의 정확도를 검증하였다. 사고발생 가능성을 내포한 운전자 분류모형을 개발한 결과 감속도($X_{deceleration}$) 및 우측방향 가속도(Y_{right}) 가 동시에 작용할 때 이 변수가 사고발생 운전자 분류의 최적 요인변수로 도출되었으며, 판별분석에 의한 예측모형은 최대 62.8%, 로지스틱회귀분석에 의한 예측모형은 최대 76.7%의 비율로 사고발생 운전자 분류가 가능한 것으로 나타났다. 또한, 모형 예측력에 대한 검증결과 84.1%의 적중률을 보이는 것으로 나타났다.

핵심어 : 디지털 운행기록, 시내버스 운전자, 가속도, 판별분석, 로지스틱회귀분석

ABSTRACT

This study aims to develop a model by which city bus drivers who are likely to cause an accident can be figured out based on the information about their actual driving records. For this purpose, from the information about the actual driving records of the drivers who have caused an accident and those who have not caused any, significance variables related to traffic accidents are drawn, and the accuracy between models is compared for the classification models developed, applying a discriminant analysis and logistic regression analysis. In addition, the developed models are applied to the data on other drivers' driving records to verify the accuracy of the models. As a result of developing a model for the classification of drivers who are likely to cause an accident, when deceleration ($X_{deceleration}$) and acceleration to the right (Y_{right}) are simultaneously in action, this variable was drawn as the optimal factor variable of the classification of drivers who had caused an accident, and the prediction model by discriminant analysis classified drivers who had caused an accident at a rate up to 62.8%, and the prediction model by logistic regression analysis could classify those who had caused an accident at a rate up to 76.7%. In addition, as a result of the verification of model predictive power of the models showed an accuracy rate of 84.1%.

Key words : Digital Tachometer Records, City Bus Drivers, Acceleration, Discriminant Analysis, Logistic Regression Analysis

* 주저자 : 명지대학교 교통공학과(박사수료)

** 공저자 및 교신저자 : 명지대학교 교통공학과 교수

† Corresponding author : Ki-jung Kum(Myongji University), E-mail kjkum@mju.ac.kr

‡ Received 11 February 2016; reviewed 15 February 2016; Accepted 16 February 2016

I. 서 론

1. 연구배경 및 목적

경찰청 통계자료에 의하면, 2013년 사업용자동차에 의한 교통사고 건수는 전체 교통사고의 22.4%를 차지하고 있으나, 비사업용자동차 등록대수 대비 4배 이상 높은 비율을 나타내고 있다. 특히 지역주민이나 학생 등 일상생활과 밀접한 교통수단이자 지역내 근거리를 연결하는 시내버스는, 전체 자동차 등록대수의 0.17%에 불과하나 교통사고는 15배 이상 많이 발생한 것으로 나타나 우려를 더하고 있다.

한편, 지난 3년(2011~2013년)간 시내버스와 관련한 교통사고 현황을 보면, “차대사람” 사고가 51.4%, “차대차” 사고가 44.0%를 차지하고 있으며, 특히, “차대사람” 사고에서는 차내전도 사고가 66.5%로 가장 많은 비중을 차지하고 있다. 또한, 법규위반 유형에 따른 사고에서는 안전운행불이행에 따른 사고가 67.4%로 전체 사고의 2/3이상을 차지하는 것으로 나타나, 시내버스 사고감소를 위해서는 사고 위험성을 내포한 운전자 관리를 위한 다양한 대책이 필요한 것으로 판단된다.

이에 본 연구에서는 시내버스 운전자의 실제 운행기록자료를 이용하여 사고발생 가능성을 내포한 운전자를 사전에 분류할 수 있는 최적 변수를 도출하고, 도출된 유의변수를 활용하여 사고발생 가능성이 높은 운전자를 추출할 수 있는 모형을 구축하였다. 이에 따라 그 결과는 운전교정 교육이나 운전자 신규 채용 시 잠재적인 위험요인을 내포한 운전자를 사전에 파악하는데 활용할 수 있으며, 결과적으로 운전자의 안전운전 환경을 도모하는데 일조할 수 있을 것으로 기대된다.

2. 연구범위 및 방법

본 연구에서는 2014년도 1월부터 6월까지 기간 중 서울, 경기, 대전지역 시내버스 운전자를 대상으로 사고를 발생했던 운전자의 사고당일 운행기록자료와 사고를 발생하지 않은 운전자의 실제 운행

기록자료를 수집·분석 하였다.

가속도 등 두 운전자 그룹간 운행특성에 차이를 나타내는 유의변수를 추출하고, 추출된 유의변수를 통하여 두 집단의 분류모형 개발 및 모형의 타당성을 검증하였다. 모형개발은 입력변수를 이용하여 목표변수를 예측하는 모형 즉, 새로운 관측치가 어떤 그룹에 속할 것인가를 결정하기 위한 여러가지 분류기법들 중 판별분석(Discriminant Analysis) 및 로지스틱회귀분석(Logistic Regression Analysis)을 활용하였다. 마지막으로 개발된 모형을 다른 운전자들의 운행기록자료에 적용하여 모형의 정확도를 검증하였다.

II. 이론적 고찰

1. 선행연구 고찰

Bagdadi and Várhelyi(2011)는 운전자 166명의 운행태를 분석한 결과 불안정한 행동이 많을수록 사고의 위험도가 높은 것으로 나타났으며, 급감속, 급핸들조작 등이 많을수록 사고와의 개연성이 높다고 밝혔다[1]. Cho(2008)는 위험상황을 자주 만드는 운전자는 그렇지 않은 운전자에 비하여 잠재적인 위험성으로 사고가 발생할 경우 그 피해가 클 가능성이 매우 높다고 하였으며[2], Cho and Lee(2007)는 2건 이상의 사고이력을 가지고 있는 그룹의 경우 무사고 그룹에 비하여 약 2배이상 위험운전을 하며, 급핸들조작 및 차선변경이 상대적으로 많다고 밝혔다[3]. Hwang(2000)은 교통사고 통계를 원인별로 분석한 결과 교통사고의 주된 요인은 운전자의 난폭운전, 법규위반 등 인적요인이 가장 많은 부분을 차지하고 있으며[4], Lee and An(2006)은 운전자의 조급한 성격으로 나타날 수 있는 급제동과 급출발이 사고로 이어질 수 있는 요인이 된다고 하였다[5]. Oh(2011)는 과속과 같은 위험운전행동 수준이 높을수록 교통사고 경험이 많으며, 과속운전과 난폭운전, 음주운전 및 주의산만 등의 운전자 의도가 개입된 위험운전행동들 간에 통계적으로 유의한 수준의 정적상관관계들이 있다고 하였다[6]. Oh and Lee(2009)는 일

반운전자에 대한 시뮬레이터 실험결과 운전자의 가·감속에 따른 위험운전과 차량핸들 조작을 통한 조정성이 복합적으로 발생할 때 가장 위험하다고 하였으며[7], Park and Lee(1997)는 사고를 잘 내지 않는 버스운전자와 사고를 많이 내는 교정교육 수강자 집단간의 차이검증 결과 성실성, 운전규칙 준수, 책임감에서 유의미한 차이가 있음을 보여주었다[8]. 한편, af Wählberg(2004)는 버스운전자의 교통사고와 가속도의 상관관계를 규명하였으며, 감속행태와 좌·우 핸들조작 행동에서 사고와 상관관계를 보였다[9]. 또한, Zegeer et al.(1993)은 사업용 차량(버스)을 대상으로 한 사고특성을 분석하여 인적요인, 차량요인 등이 사고와 관련이 있음을 보였으며, 특히 전방주시태만, 안전거리 미확보 등에 의한 추돌 및 측면충돌 사고가 많아 사업용 차량 운전자에 대한 교육의 필요성을 강조하였다[10].

2. 연구의 차별성

지금까지 많은 연구자들이 운전자의 성격 및 태도와 같은 심리적 특성을 활용하여 교통사고의 위험성을 높이는 운전자의 인적요인을 찾아내기 위하여 노력하여 왔으나, 대부분의 선행연구들은 다음 두가지 형태로 구분할 수 있다.

첫 번째, 기존 많은 연구들은 운전자의 위험요인을 도출하고 평가체계를 갖추기 위하여 운전자의 성격 및 태도와 관련된 사항에 대하여 주로 설문조사의 방법을 취하였으며, 이를 바탕으로 교통사고 발생을 야기하는 운전자 요인들의 영향을 추정하는 모형을 구축하고 검증을 실시하였다.

두 번째, 위험운전행태를 분류하고 설명할 수 있는 기준을 마련하기 위하여 실제 운행자료가 아닌 실험용 차량을 이용한 실험환경을 조성하였다. 조성된 실험환경하에서의 주행시뮬레이터 분석을 통하여 수집된 자료를 활용하여 위험운전에 관한 연구를 진행하였다. 이처럼 기존의 교통사고와 관련된 운전자의 인적요인 도출에 관한 연구들은 설문조사와 실험을 통한 연구가 대부분을 차지하였으며, 교통사고 발생자료 또는 경험적 자료를 활용하

여 사고발생과 운전자 인적요인과의 관계를 분석하여 왔다. 하지만 운행정보 데이터 수집의 한계성 및 접근성 제한 등으로 연구는 극히 제한적일 수 밖에 없으며, 특히 대중교통수단으로써 운송을 목적으로 하는 사업용자동차에 대한 연구는 미미한 실정이다. 본 연구가 갖는 차별성은, 기존 많은 연구가 운전자의 성격 및 태도 즉 운전자의 심리적 요인과 교통사고와의 관계에 주목했다면, 본 연구는 운전자의 행태적인 측면이 주요 관심사항이 된다고 할 수 있다. 조성된 실험조건이 아닌 실제 도로상에서 발생하는 차량들의 운행정보를 활용하기 때문에 실질적 행태를 잘 반영한다고 할 수 있다. 또한 본 연구에서는 사업용 운전자의 실제 운행자료를 바탕으로 사고운전자와 비사고 운전자를 구분할 수 있는 요인에 대해 연구하였다. 두 운전자 집단의 유의한 특성변수를 대상으로 모형내 변수의 영향력 및 통계적 유의성이 확보되지 않은 변수들을 순차적으로 제거하여 두 집단의 분류요인에 대해 가장 잘 설명할 수 있는 모형을 구축하였다.

연구결과, 사고자와 비사고자를 구분할 수 있는 기준 제시를 통해 교통사고 발생 가능성이 있는 운전자를 사전에 파악하여 운전교정 교육 등 사고 발생 위험군 운전자관리에 활용될 수 있을 것으로 판단되며, 나아가 타 업종 운전자를 대상으로 한 사고 위험군 운전자를 판별할 수 있는 연구의 기초로 활용될 수 있을 것이다.

III. 분석용 자료구축 및 분석방법론

1. 분석용 자료구축

본 연구에서는 2014년도 1월부터 6월까지 기간 중 서울, 경기, 대전지역 시내버스 운전자를 연구대상으로 설정하였으며, 동 기간 중 사고를 발생한 운전자와 사고를 발생하지 않은 운전자의 운행기록 자료를 수집하여 분석에 활용하였다. 수집된 자료중 1~2월의 자료는 모형개발을 위하여 활용하고 나머지 3~6월의 자료는 개발된 모형의 검증을 위하여 사용하였다. 특히, 무사고자는 월별, 지역별 사고자

구성비율을 고려하여 사고자의 2배수 수준으로 자료를 수집하였다.

연구대상 총 표본 및 자료의 세부 구성내역은 <Table 1> ~ <Table 3>과 같다.

<Table 1> Number of Samples of Research Subjects

Classification	Drivers who have caused an accident	Drivers who have not caused an accident	Total
For developing models	43	86	129
For verifying models	63	63	126
Total	106	149	255

<Table 2> Composition of Samples of Research Subjects by Region

Month Class.	January		February		Total		
	No. of data	Comp onent ratio	No. of data	Comp onent ratio	No. of data	Comp onent ratio	
w/ accident	Seoul	3	7.0	6	14.0	9	20.9
	Daejeon	8	18.9	7	16.3	15	34.9
	Gyeonggi	13	30.2	6	14.0	19	44.2
	Subtotal	24	55.8	19	44.2	43	100
w/o accident	Seoul	6	7.0	12	14.0	18	20.9
	Daejeon	16	18.9	14	16.3	30	34.9
	Gyeonggi	26	30.2	12	14.0	38	44.2
	Subtotal	48	55.8	38	44.2	86	100
Total	72	55.8	57	44.2	129	100	

<Table 3> Composition of Samples for Verification

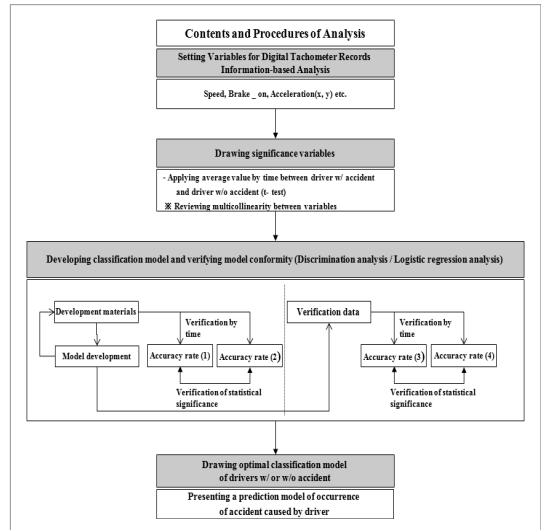
Class.	Drivers who have caused an accident	Drivers who have not caused an accident	Total
Seoul	19	40	59
Daejeon	16	12	28
Gyeonggi	28	11	39
Total	63	63	126

2. 분석방법론

본 연구는 속도, 브레이크 작동횟수 및 가속도 등 운행기록 자료로부터 수집된 운전자 특성변수가 각 운전자의 사고발생 가능성을 얼마나 정확하게

예측하느냐에 중점을 두고 있다. 이를 위하여 본 연구에서는, 종속변수를 예측하기 위한 통계적 기법 중 대표적인 로지스틱회귀분석(Logistic Regression Analysis)과 판별분석(Discriminant Analysis)을 사용하였으며, 각 분석기법은 자료의 특성에 따라 다른 효율성과 정확성을 나타내기 때문에 두 분류기법의 비교를 통하여 최적의 분류모형을 제시하였다.

시내버스 운전자의 사고예측모형 개발을 위한 구체적인 분석내용 및 절차는 <Fig. 1>과 같다.



<Fig. 1> Procedures and Methods of Analysis

첫째, 운행기록계에서 수집된 자료중 교통사고 영향요인과 관련된 분석용 변수를 도출하고, 사고자와 비사고자 집단간 유의한 설명변수를 추출하였다. 둘째, 추출된 설명변수를 모형개발용 자료에 적용하여 투입된 변수의 영향력이 적은 변수 및 통계적으로 유의하지 않은 변수를 하나씩 제거하면서 모형을 개발하고, 모형간 분류정확도 비교분석을 통하여 분류 정확도가 가장 높은 모형을 최적 모형으로 선택하였다. 또한, 개발된 모형은 1일 전체 운행시간 중 임의 1시간에 대한 자료를 바탕으로 개발되었기 때문에 임의 추출된 다른 시간대에 모형을 적용하여 모형의 분류 정확도 및 예측 분류된 집단에 대한 통계적 검증을 실시하였다.

셋째, 모형개발 단계와 동일한 방법으로, 개발된 최적 모형을 검증용 자료에 적용하여 모형의 정확도를 검증하고, 또한 임의 다른 시간대에 모형을 적용하여 시간대별 모형 적용의 타당성을 검증하였다.

1) 판별분석(Discriminant Analysis)

판별분석은 자료를 분석한 근거를 가지고 의사를 결정해야 할 때 널리 사용된다. 또한, 여러 개의 독립변수가 지닌 특성들을 파악하여 그 표본이 어느 집단에 속하는가를 판별하고, 분석을 통해 미지의 사례가 어떤 집단에 속할 것인지를 분석할 수 있을 뿐만 아니라, 어떤 변수가 종속변수에 가장 큰 영향을 미칠 것인가도 알 수 있다.

판별함수는 각 독립변수와 이에 상응하는 가중치를 곱하여 이를 모두 합해 각 표본단위의 판별점수를 구하고, 이를 통합하여 집단의 평균값을 구한다. 또한 판별계수는 회귀계수와 같이 각 독립변수의 상대적 중요도를 의미하며, 판별함수는 집단구분의 예측력을 의미하기 때문에 판별함수가 좋다는 것은 곧 독립변수들이 집단을 잘 구분한다는 것을 의미한다. 일반적인 판별함수식은 다음과 같이 나타낼 수 있다.

$$Z = WX = B_0 + W_1X_1 + W_2X_2 + \dots + W_pX_p \quad (1)$$

Z : 판별방정식(Discriminant Equation)

B₀ : 판별상수(Discriminant Score)

W₁ ~ W_p : 판별함수 계수

X₁ ~ X_p : 유의변수

2) 로지스틱회귀분석(Logistic Regression Analysis)

로지스틱회귀분석은 어떤 사건이 일어날 확률을 추정하기 위해 고안된 통계기법으로, 종속변수가 범주형인 자료를 분석하는데 적합하다. 로지스틱회귀분석은 선형회귀분석이 갖는 한계에서부터 출발하며 선형회귀분석이 주어진 독립변수로부터 종속변수의 평균값을 예측하는 것이라고 한다면, 로지스틱회귀분석은 주어진 독립변수로부터 종속변수의 범주 중 어느 하나의 비율을 예측하는 것이라고 할 수 있다.

하나의 독립변수를 고려한 로지스틱 회귀모형은

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} \quad (2)$$

로 표현되며, p개의 설명변수를 고려하면

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (3)$$

와 같이 확장 가능하다. 한편, 식(2)와 식(3)의 좌변에 있는 로짓(Logit)은 아래와 같이 표기되기도 한다.

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) \quad (4)$$

그리고 식(3)의 모형으로부터 확률은 아래와 같이 계산된다.

$$p_i = \frac{\exp(x'_i \beta)}{1 + \exp(x'_i \beta)} = \frac{1}{1 + \exp(-x'_i \beta)} \quad (5)$$

IV. 모형개발 및 검증

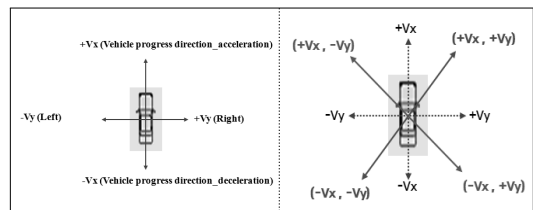
1. 유의변수 선정

1) 분석용 변수의 구성

운행기록계에서 수집된 항목 중 차량속도(v), 엔진 회전수(rpm) 등 5개 항목을 분석에 사용하였으며, 가속도의 경우 가감속 및 좌우 방향성을 고려하여 변수를 추가로 생성하였다.

<Table 4> Variables Used for Analysis

Item	Unit	Definition of variable
Vehicle speed	km/h	Vehicle speed
No. of engine rotation per minute	rpm	Engine RPM
Brake signal	1, 0	Brake operation
Acceleration	(±)ΔVx	Unit time acceleration
	(±)ΔVy	Unit time acceleration



<Fig. 2> Definition of Acceleration Variable

<Table 5> Processing Acceleration Variables

Item	Unit	Definition of variable
Resultant acceleration	X, Y	m/sec^2 $\Delta V_x, \Delta V_y$ Resultant force
	+X, +Y	m/sec^2 $+\Delta V_x, +\Delta V_y$ Resultant force
	+X, -Y	m/sec^2 $+\Delta V_x, -\Delta V_y$ Resultant force
	-X, +Y	m/sec^2 $-\Delta V_x, +\Delta V_y$ Resultant force
	-X, -Y	m/sec^2 $-\Delta V_x, -\Delta V_y$ Resultant force
Lateral acceleration (Y)	m/sec^2	$ \pm \Delta V_y $

이상과 같이 분석에 사용될 최종변수를 선정하였으며, 최종 분석용 변수는 <Table 6>과 같다.

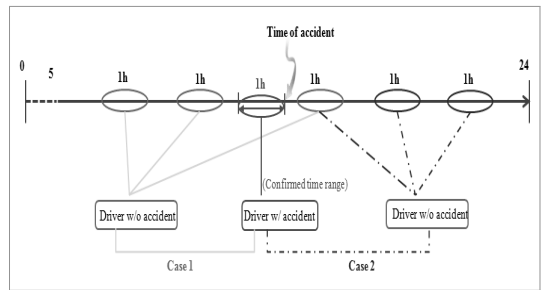
<Table 6> List of Variables for Analysis

No	Variable name	Description of variable	Concept applied
1	V	Velocity	
2	rpm	No. of engine rotation	
3	brake_on	Brake operation signal	
4	Vx	Acceleration ($\pm V_x$)	Acceleration in the direction of progress of vehicle
5	Xa ^d	Acceleration (+Vx)	Acceleration
6	Xd ^d	Acceleration (-Vx)	Acceleration
7	Vy	Acceleration ($\pm V_y$)	(Left/Right) Lateral acceleration
8	Yr ^d	Acceleration (+Vy)	Acceleration in action to the right
9	Yl ^d	Acceleration (-Vy)	Acceleration in action to the left
10	Y	Size of acceleration ($\pm V_y$)	Size of (Left/Right) lateral acceleration
11	XY	Resultant acceleration ($\pm V_x, \pm V_y$)	Size of resultant acceleration
12	XaYr	Resultant acceleration (+Vx, +Vy)	Reduction + acceleration to the right
13	XaYl	Resultant acceleration (+Vx, -Vy)	Reduction + acceleration to the left
14	XdYr	Resultant acceleration (-Vx, +Vy)	Reduction + acceleration to the right
15	XdYl	Resultant acceleration (-Vx, -Vy)	Reduction + acceleration to the left

Note) 1. a : acceleration d : deceleration r : right l : left

2) 유의변수 선정

유의변수 선정을 위하여 사고발생 운전자 및 사고 미발생 운전자의 1일 운행자료 중 각각 1시간 자료를 대상으로 t-검정을 수행하였다. 사고발생 운전자의 경우 사고발생시점을 기준으로 사고전 1시간 자료를 추출하고, 사고 미발생 운전자의 경우 05시~24시 중 무작위로 1시간을 추출하여 집단간 변수의 유의성을 검정하고 동 과정을 20회 반복 실시하여 시간대별로 유의변수 선정의 타당성을 검증하였다. 유의변수 선정을 위한 방법론을 도식화하면 <Fig. 3>과 같다.



<Fig. 3> Scenario of Selection of Significance Variable

먼저, case 1을 대상으로 운행기록계 정보에서 설명변수로 추출된 차량속도(v), 브레이크 작동횟수, 가속도 등 5개 변수와 추가로 생성된 가속도 관련 10개 변수를 대상으로 두 집단간 차이 유무를 검정하였다.

사고발생 운전자 및 사고 미발생 운전자 그룹간 설명변수의 유의성 검정결과 Vy, XY, Y 3개 항목이 신뢰수준 95% ($\alpha = 0.05$)에서 그룹간 차이가 유의한 것으로 나타났으며, Xd, XdYr 및 XdYl 항목이 신뢰수준 99% ($\alpha = 0.01$)에서 그룹간 차이가 유의한 것으로 나타났다. 반면 속도(v) 및 브레이크 작동회수, 진행방향 가속도(Vx) 등 9개 항목은 그룹간 차이가 유의하지 않은 것으로 분석되었다. case 1의 경우와 같이 case 2~case 20까지의 유의변수 검정 과정을 반복 수행한 결과, Vy, XY, Y 3개 항목은 신뢰수준 95% ($\alpha = 0.05$)에서, Xd, XdYr 및 XdYl 3개

항목은 신뢰수준 99% ($\alpha = 0.01$)에서 그룹간 차이가 유의한 것으로 나타났으며, 속도(v) 및 브레이크 작동회수, 진행방향 가속도(Vx) 등 9개 항목은 그룹간 차이가 유의하지 않은 것으로 분석되었다. t-검정 결과에 따른 최종 유의변수 선정결과는 <Table 8>과 같다.

<Table 7> Result of Significance Test on Descriptive Variables between Groups (Case 1)

Variable	Classification of group (1 : w/ accident 0 : w/o accident)	Average	Standard deviation	t-value	p
V	1	19.190	9.311	-1.446	0.151
	0	22.105	11.456		
brake_on	1	1349.558	756.425	1.291	0.199
	0	1195.686	570.714		
rpm	1	776.678	268.641	-0.889	0.378
	0	815.596	142.505		
Vx	1	0.094	0.283	0.120	0.905
	0	0.089	0.201		
Vy	1	0.003	0.153	2.268	0.025*
	0	-0.039	0.054		
XY	1	0.243	0.254	2.269	0.026*
	0	0.143	0.195		
Xa	1	0.163	0.223	1.412	0.160
	0	0.109	0.196		
Xd	1	-0.149	0.141	-3.122	0.003**
	0	-0.078	0.068		
Yr	1	0.078	0.132	1.718	0.092
	0	0.043	0.042		
Yl	1	-0.079	0.048	-1.083	0.281
	0	-0.070	0.048		
Y	1	0.095	0.131	2.309	0.023*
	0	0.058	0.053		
XaYr	1	0.186	0.252	1.888	0.063
	0	0.103	0.196		
XaYl	1	0.182	0.219	1.462	0.146
	0	0.126	0.198		
XdYr	1	0.132	0.174	4.185	0.000**
	0	0.021	0.024		
XdYl	1	0.135	0.125	5.121	0.000**
	0	0.054	0.056		

Note) *: Significant at a level of p-value 0.05,
 **: Significant at a level of p-value 0.01

<Table 8> Result of Selection of Significance Variables

Variable Class	Vy	XY	Xd	Y	XdYr	XdYl
case1	*	*	**	*	**	**
case2	*	*	**	*	**	**
case3	*	*	**	*	**	**
case4	*	*	**	*	**	**
case5	*	*	**	*	**	**
case6	*	*	**	*	**	**
case7	*	*	**	*	**	**
case8	*	*	**	*	**	**
case9	*	*	**	*	**	**
case10	*	*	**	*	**	**
case11	*	*	**	*	**	**
case12	*	*	**	*	**	**
case13	*	*	**	*	**	**
case14	*	*	**	*	**	**
case15	*	*	**	*	**	**
case16	*	*	**	*	**	**
case17	*	*	**	*	**	**
case18	*	*	**	*	**	**
case19	*	*	**	*	**	**
case20	*	*	**	*	**	**

Note) *: Significant at a level of p-value 0.05,
 **: Significant at a level of p-value 0.01

3) 다중공선성

t-검정 결과 유의한 변수로 선정된 변수들간의 연관성을 파악하기 위하여 다중공선성을 확인하였다. 기본적으로 다중공선성 진단은 공차한계가 0.1 이하로 작거나 VIF(Variance Inflation Factor) 가 10 이상이면 다중공선성의 문제가 있다고 본다.

<Table 9> Result of Analysis of Multicollinearity

Variable	Tolerance limit	VIF
Vy	0.076	13.109
XY	0.586	1.706
Y	0.402	2.486
XdYl	0.067	14.888
Xd	0.370	2.705
XdYr	0.047	21.496

Vy, Y, XY, Xd, XdYr 및 XdYl 등 6개 가속도 변수에 대한 다중공선성 분석 결과, Vy, XdYr 및

XdY1 3개 변수의 공차한계가 0.047~0.076으로 0.1 이하이고 VIF도 13.109~21.496으로 10이상으로 높기 때문에 다중공선성을 의심해 볼 수 있다. 다중공선성 문제 해결을 위해서는 표본수를 확대하거나 해당변수를 제거하여야 하지만, 본 연구에서는 중요 설명변수가 제거될 가능성이 있기 때문에 유의변수 선정과정에서 다중공선성 분석 결과를 적용하지는 않고 사고 운전자 분류모형 개발 단계에서 다중공선성 의심변수를 제거한 모형을 개발하였다.

2. 모형개발 결과

1) 판별분석에 의한 모형개발

먼저 V_y , Y , XY , X_d , X_dY_r 및 X_dY_1 등 6개 가속도 변수 중 다중공선성 의심변수 V_y 를 제외한 나머지 5개 변수를 사용하여 판별분석을 실시한 결과, 고유값(Eigenvalue) 0.322, 정준상관계수 0.493, Wilk's Lambda 값 $\lambda = 0.757(p < 0.001)$ 으로 나타났으며 Box's M 검정에 의한 공분산행렬의 동질성에 대한 검정에서는 통계적 유의값이 0.000으로 나타나, 모집단 공분산행렬이 동일하다는 영가설을 만족하지는 못하였다.

정준판별함수의 계수를 보면, 사고 발생 운전자와 사고 미발생 운전자의 분류 변인으로서 가장 영향력이 큰 변수는 X_dY_r (감속+우측방향 작용 가속도), 영향력이 가장 낮은 변수는 X_d (감속도)임을 알 수 있다. 판별함수를 통하여 두 집단에 대한 분류 정확도를 분석한 결과, 사고 미발생 운전자 집단을 사고 미발생 운전자 집단으로 정확하게 분류하는 비율은 96.5%, 사고발생 운전자 집단을 사고발생 운전자 집단으로 분류하는 비율은 62.8%였으며, 사고발생 운전자 집단을 사고 미발생 운전자 집단으로 잘못 예측하는 오류비율은 37.2%로 나타났다. 위 결과를 토대로 모형내 변수의 영향력이 가장 적은 변수를 차례로 제거해 가면서 추가 모형을 개발하는 한편, 동일한 방법으로 다중공선성 의심변수를 변경하여 판별모형을 개발하였다.

다중공선성 의심변수 V_y 를 제거한 후 나머지 5개 유의변수를 판별분석에 적용한 경우에는 사고

미발생 운전자를 96.5% 정분류 하고, 사고발생 운전자에 대해서는 62.8% 수준의 분류율을 나타내었으며, 모형내 영향력이 가장 낮은 변수를 하나씩 제거하여 모형의 분류정확도를 분석한 결과, 영향력이 가장 높은 유의변수는 X_dY_r 이었으나 사고발생 운전자의 분류율은 단일 변수에 의해서는 48.8%로 가장 낮았고, 다른 유의변수와 함께 모형개발에 투입시 분류율이 가장 높게 나오는 것으로 분석되었다.

다중공선성 의심변수 X_dY_r 를 제거한 후 나머지 유의변수를 판별분석에 적용한 경우에는 사고 미발생 운전자의 분류율은 98.8%, 사고발생 운전자에 대해서는 53.5%의 분류율을 나타내었다. 한편, 모형내 영향력이 가장 낮은 변수를 하나씩 제거하여 모형의 분류정확도를 분석한 결과, 가장 영향력이 높은 유의변수는 X_dY_1 로 사고발생 운전자의 분류율은 X_dY_1 단일 변수를 모형개발에 투입했을 때 62.8%로 가장 높게 나오는 것으로 분석되었다.

다중공선성 의심변수 X_dY_1 를 제거한 후 나머지 유의변수를 판별분석에 적용한 경우에는 사고 미발생 운전자의 분류율은 96.5%, 사고발생 운전자에 대해서는 55.8%의 분류율을 나타내었다.

또한, 모형내 영향력이 가장 낮은 변수를 하나씩 제거하여 모형의 분류정확도를 분석한 결과, 가장 영향력이 높은 유의변수는 X_dY_r 로 나타났으나 X_dY_r 단일 변수를 모형개발에 투입했을 때 사고발생 운전자의 분류율은 48.8%로 가장 낮게 나오는 것으로 분석되었다.

2) 로지스틱회귀분석에 의한 모형개발

6개 가속도 변수 중 다중공선성 의심변수 V_y 를 제외하고 나머지 5개 변수를 모두 진입시켜 사고발생 운전자 및 사고 미발생 운전자 그룹에 대한 로지스틱회귀분석을 실시한 결과, 모형의 적합도 검증에서는 -2LL(-2Log Likelihood) 값의 변화를 통한 χ^2 통계량이 신뢰수준 95%($\alpha = 0.05$)에서 유의한 것으로 나타났으며, 종속변수의 실제값과 예측된 값의 적합성 여부를 측정하는 Hosmer&Lemeshow Test의 p 값이 0.435($p > 0.05$)로 귀무가설을 채택하여 모형은 적합한 것으로 나타났다.

<Table 10> Result of Discriminant Analysis

Multicollinearity variable (exclusion variable)	Input factor	Eigen value	Canonical correlation coefficient	Wilk's Lambda	Discriminant functional formula	Classification accuracy(%)		
						w/o accident	w/ accident	Total
Vy	Y, XY, Xd, XdYr, XdYl	0.322	0.493	0.757	$Z = -0.760 + 4.558(XdYl) + 7.538(XdYr) - 2.068(Y) + 0.023(Xd) + 0.591(XY)$	96.5	62.8	85.3
	Y, XY, XdYr, XdYl	0.322	0.493	0.757	$Z = -0.762 + 4.547(XdYl) + 7.526(XdYr) - 2.057(Y) + 0.593(XY)$	96.5	62.8	85.3
	Y, XdYr, XdYl	0.318	0.491	0.759	$Z = -0.744 + 4.956(XdYl) + 7.417(XdYr) - 1.212(Y)$	96.5	62.8	85.3
	XdYr, XdYl	0.316	0.490	0.760	$Z = -0.814 + 5.215(XdYl) + 6.781(XdYr)$	94.2	62.8	83.7
	XdYr	0.270	0.461	0.787	$Z = -0.567 + 9.808(XdYr)$	95.3	48.8	79.8
XdYr	Vy, XY, Xd, Y, XdYl	0.353	0.511	0.739	$Z = -0.649 + 12.520(XdYl) - 2.287(Y) + 0.787(Xd) + 0.417(XY) + 7.819(Vy)$	98.8	53.5	83.7
	Vy, Xd, Y, XdYl	0.351	0.510	0.740	$Z = -0.613 + 12.987(XdYl) - 1.864(Y) + 1.059(Xd) + 7.911(Vy)$	98.8	53.5	83.7
	Vy, Y, XdYl	0.350	0.509	0.741	$Z = -0.697 + 12.028(XdYl) - 1.286(Y) + 7.404(Vy)$	98.8	53.5	83.7
	Vy, XdYl	0.346	0.507	0.743	$Z = -0.772 + 11.662(XdYl) + 6.832(Vy)$	97.7	58.1	84.5
	XdYl	0.207	0.414	0.829	$Z = -0.945 + 11.701(XdYl)$	73.3	62.8	69.8
XdYl	Vy, XY, Xd, Y, XdYr	0.302	0.481	0.768	$Z = -0.766 + 10.288(XdYr) - 2.343(Y) - 1.226(Xd) + 0.954(XY) - 1.912(Vy)$	96.5	55.8	82.9
	Vy, Xd, Y, XdYr	0.293	0.476	0.773	$Z = -0.694 + 11.109(XdYr) - 1.307(Y) - 0.787(Xd) - 2.547(Vy)$	95.3	58.1	82.9
	Vy, Y, XdYr	0.293	0.476	0.774	$Z = -0.629 + 11.850(XdYr) - 1.789(Y) - 2.797(Vy)$	95.3	55.8	82.2
	Vy, XdYr	0.288	0.473	0.777	$Z = -0.729 + 11.272(XdYr) - 3.084(Vy)$	94.2	53.5	80.6
	XdYr	0.270	0.461	0.787	$Z = -0.567 + 9.808(XdYr)$	95.3	48.8	79.8

회귀계수를 보면, 사고 발생 운전자와 사고 미발생 운전자의 분류변인으로서 가장 영향력이 있는 변수는 XdYr(감속+우측방향 작용 가속도)로서 값이 커질수록 사고발생 운전자로 분류될 가능성이 더 커지며, 반면 큰 영향력을 가지는 Y(좌우 횡방향 가속도 크기)의 경우 값이 커질수록 사고발생 운전자로 분류될 가능성이 낮아짐을 알 수 있다. 분석결과에 따른 분류 정확도를 분석한 결과 사고 미발생 운전자 집단을 사고 미발생 운전자 집단으로 정확하게 분류하는 비율은 94.2%, 사고발생 운전자 집단을 사고발생 운전자 집단으로 분류하는 비율은 79.1%로 나타났으며, 전반적인 분류 정확도는 89.1%로 나타났다. 최적 분류모형 개발을 위하여 통계적으로 유의하지 않은 변수를 제거하고 영향력이 가장 적은 변수부터 모형에서 제거해 가면서 개발된 모형의 분류정확도 변화량을 분석하는 한편, 동일한 방법으로 다중공선성 의심변수를 변경하여 분류모형

을 개발하였다.

변수의 영향력과 통계적 유의성을 모두 고려한 결과, Y, XY, XdYr 및 XdYl 4개 변수를 모형개발에 투입한 경우 사고 미발생 운전자 94.2%, 사고발생 운전자 74.4%의 분류율을 나타내었으며, 사고발생 운전자의 분류율에 대한 정확도가 가장 높은 경우는 XdYr 단독변수를 모형개발에 투입하는 경우에 76.7%의 분류율을 나타내었다.

다중공선성 의심변수 XdYr 및 통계적으로 유의하지 않은 설명변수를 제거한 경우에는, Vy, Y, XdYl 3개 유의변수가 적용된 모형에서 전체 분류 정확도가 87.6%로 가장 높게 나타났으며, 사고발생 운전자 분류 정확도 또한 69.8%로 가장 높게 나타났다. 다중공선성 의심변수 XdYl 및 통계적으로 유의하지 않은 설명변수를 제거한 경우에는 XdYr 단독변수를 모형개발에 적용한 경우, 사고 미발생 운전자 및 사고발생 운전자에 대한 분류정확도는 각

<Table 11> Result of Logistic Regression Analysis

Multicollinearity variable (exclusion variable)	Input factor	-2Log Likelihood	Model Chi-square (p-value)	Pseudo R ²			Hosmer& Lemeshow Test	Model formula	Classification accuracy(%)		
				Cox& Snell	Nagelkerke	McFadden			w/o accident	w/ accident	Total
Vy	XY, Xd, Y, XdYr, XdYl	72.532	91.688 (0.000)	0.509	0.707	0.558	0.435	Logit=-3.071+39.325(XdYl)+59.784(XdYr)-50.161(Y)+7.597(Xd)+4.416(XY)	94.2	79.1	89.1
	XY, Y, XdYr, XdYl	74.227	89.993 (0.000)	0.502	0.698	0.548	0.722	Logit=-3.717+37.935(XdYl)+52.133(XdYr)-46.421(Y)+5.078(XY)	94.2	74.4	87.6
	Y, XdYr, XdYl	85.905	78.316 (0.000)	0.455	0.632	0.477	0.032	Logit=-2.957+28.357(XdYl)+41.014(XdYr)-26.289(Y)	93.0	74.4	86.8
	XdYr, XdYl	92.144	72.077 (0.000)	0.428	0.595	0.439	0.875	Logit=-3.418+7.946(XdYl)+44.227(XdYr)	90.7	74.4	85.3
	XdYr	94.110	70.111 (0.000)	0.419	0.582	0.427	0.172	Logit=-2.991+49.062(XdYr)	91.9	76.7	86.8
XdYr	Vy, XY, Xd, Y, XdYl	79.689	84.532 (0.000)	0.481	0.668	0.515	0.857	Logit=-2.898+92.660(XdYl)-40.232(Y)+6.687(Xd)+2.883(XY)+50.110(Vy)	96.5	67.4	86.8
	Vy, XY, Y, XdYl	81.040	83.180 (0.000)	0.475	0.660	0.507	0.341	Logit=-3.437+82.530(XdYl)-37.226(Y)+3.535(XY)+42.169(Vy)	96.5	67.4	86.8
	Vy, Y, XdYl	87.848	76.373 (0.000)	0.447	0.621	0.465	0.767	Logit=-2.920+69.498(XdYl)-25.315(Y)+35.551(Vy)	96.5	69.8	87.6
	Vy, XdYl	91.259	79.962 (0.000)	0.432	0.600	0.444	0.958	Logit=-3.343+54.869(XdYl)+9.785(Vy)	94.2	67.4	85.3
	XdYl	133.723	30.498 (0.000)	0.211	0.292	0.186	0.026	Logit=-2.129+16.973(XdYl)	86.0	44.2	72.1
XdYl	Vy, XY, Xd, Y, XdYr	82.526	81.694 (0.000)	0.469	0.652	0.497	0.414	Logit=-3.028+74.626(XdYr)-26.370(Y)+4.626(Xd)+4.279(XY)-11.145(Vy)	91.9	79.1	87.6
	Vy, XY, Y, XdYr	83.354	80.867 (0.000)	0.466	0.647	0.492	0.798	Logit=-3.430+69.491(XdYr)-25.895(Y)+4.694(XY)-12.661(Vy)	91.9	74.4	86.0
	XY, Y, XdYr	84.976	79.244 (0.000)	0.459	0.637	0.483	0.513	Logit=-3.549+60.867(XdYr)-11.277(Y)+4.386(XY)	91.9	79.1	87.6
	XY, XdYr	87.163	77.058 (0.000)	0.450	0.625	0.469	0.560	Logit=-4.004+56.385(XdYr)+3.558(XY)	91.9	72.1	85.3
	XdYr	94.110	70.111 (0.000)	0.419	0.582	0.427	0.172	Logit=-2.991+49.062(XdYr)	91.9	76.7	86.8

각 91.9%, 76.7%로 변수의 통계적 유의성을 고려할 경우 가장 높은 결과가 도출되었다.

분석방법에 따른 사고발생 운전자 분류정확도를 비교해 보면, 변수의 통계적 유의성 및 모형내 영향력을 고려시 판별분석에 의하여 최대 62.8%, 로지스틱회귀분석에 의하여 최대 76.7%의 분류 정확도

를 나타내었다. 모형의 적합도와 관련한 Pseudo R2 값은 Hosmer&Lemeshow 검정결과 p<0.05인 경우를 제외하고 모든 경우에서 Cox&Snell 0.419~0.509, Nagelkerke 0.582~0.707, McFadden 0.427~0.558로 나타나 모형의 설명력은 양호한 것으로 판단된다.

모형개발 결과를 정리하면 <Table 12>와 같다.

<Table 12> Comparison of Results of Development of Models According to Method of Analysis

Multicollinearity variable (exclusion variable)	Input factor	Discriminant analysis		Variable	Logistic regression analysis	
		Accident	Classification accuracy %		Accident	Classification accuracy %
Vy	XY, Xd, Y, XdYr, XdYl	No occurrence Occurrence Total %	96.5 62.8 85.3	XY, Xd*, Y, XdYr, XdYl	No occurrence Occurrence Total %	94.2 79.1 89.1
	XY, Y, XdYr, XdYl	No occurrence Occurrence Total %	96.5 62.8 85.3	XY, Y, XdYr, XdYl	No occurrence Occurrence Total %	94.2 74.4 87.6
	Y, XdYr, XdYl	No occurrence Occurrence Total %	96.5 62.8 85.3	Y, XdYr, XdYl	No occurrence Occurrence Total %	93.0 74.4 86.8
	XdYr, XdYl	No occurrence Occurrence Total %	94.2 62.8 83.7	XdYr, XdYl*	No occurrence Occurrence Total %	90.7 74.4 85.3
	XdYr	No occurrence Occurrence Total %	95.3 48.8 79.8	XdYr	No occurrence Occurrence Total %	91.9 76.7 86.8
XdYr	Vy, XY, Xd, Y, XdYl	No occurrence Occurrence Total %	98.8 53.5 83.7	Vy, XY, Xd*, Y, XdYl	No occurrence Occurrence Total %	96.5 67.4 86.8
	Vy, Xd, Y, XdYl	No occurrence Occurrence Total %	98.8 53.5 83.7	Vy, XY, Y, XdYl	No occurrence Occurrence Total %	96.5 67.4 86.8
	Vy, Y, XdYl	No occurrence Occurrence Total %	98.8 53.5 83.7	Vy, Y, XdYl	No occurrence Occurrence Total %	96.5 69.8 87.6
	Vy, XdYl	No occurrence Occurrence Total %	97.7 58.1 84.5	Vy, XdYl	No occurrence Occurrence Total %	94.2 67.4 85.3
	XdYl	No occurrence Occurrence Total %	73.3 62.8 69.8	XdYl	No occurrence Occurrence Total %	86.0 44.2 72.1
XdYl	Vy, XY, Xd, Y, XdYr	No occurrence Occurrence Total %	96.5 55.8 82.9	Vy*, XY, Xd*, Y*, XdYr	No occurrence Occurrence Total %	91.9 79.1 87.6
	Vy, Xd, Y, XdYr	No occurrence Occurrence Total %	95.3 58.1 82.9	Vy*, XY, Y*, XdYr	No occurrence Occurrence Total %	91.9 74.4 86.0
	Vy, Y, XdYr	No occurrence Occurrence Total %	95.3 55.8 82.2	XY, Y*, XdYr	No occurrence Occurrence Total %	91.9 79.1 87.6
	Vy, XdYr	No occurrence Occurrence Total %	94.2 53.5 80.6	XY, XdYr	No occurrence Occurrence Total %	91.9 72.1 85.3
	XdYr	No occurrence Occurrence Total %	95.3 48.8 79.8	XdYr	No occurrence Occurrence Total %	91.9 76.7 86.8

Note) * : $p > 0.05$

3) 모형개발 타당성 검증

모형개발은 1일 전체 운행시간 중 임의 1시간에 대한 자료를 바탕으로 하였기 때문에 전체 운행시간대에 대한 검증, 즉 임의 추출된 다른 시간대에 적용하여 모형의 적중률 및 예측 분류된 집단에 대한 통계적 검증을 수행하였다. 모형에 대한 타당성 검증은 개발된 모형 중 사고발생 운전자의 분류 정

확도가 가장 높은 모형을 대상으로 하였으며, <Table 12>의 결과에서 나타난 XdYr 단독 변수에 의한 분류모형을 대상으로 하였다. 무작위로 임의 시간대(t2)를 추출하고 추출된 시간대에 모형을 적용하여 적중률을 산출한 결과 사고 미발생 운전자의 경우 90.7%, 사고 발생 운전자의 경우 76.7%의 적중률을 보였다. 또한, 예측 분류된 집단의 시간대

별 동질성에 대한 통계적 검증에서는 t-검정 결과 신뢰수준 95%($\alpha=0.05$)에서 시간대별로 차이는 유의하지 않은 것으로 분석되었다.

<Table 13> Test on Accuracy Rate of Model (t1~t2)

Accuracy rate (I) (t1)			Accuracy rate (II) (t2)				
Observation	Predictive value		Classification accuracy %	Observation	Predictive value		Classification accuracy %
	No occurrence	Occurrence			No occurrence	Occurrence	
No occurrence	79	7	91.9	No occurrence	78	8	90.7
Occurrence	10	33	76.7	Occurrence	10	33	76.7
Total			86.8	Total			86.0

<Table 14> Test on Homogeneity of Predictive Classification Group (t1~t2)

Verification variable	Predictive value		N	Average	Standard deviation	t	Significance probability
	No occurrence	Accuracy rate (I) (t1)					
XdYr	No occurrence	Accuracy rate (I) (t1)	89	0.016	0.018	-0.098	0.922
		Accuracy rate (II) (t2)	88	0.017	0.018		
	Occurrence	Accuracy rate (I) (t1)	40	0.149	0.173	0.051	0.959
		Accuracy rate (II) (t2)	41	0.147	0.171		

3. 모형검증

개발된 모형의 검증을 위하여 시내버스 운전자 운행기록자료 중 모형개발에 사용하지 않은 126명의 자료를 활용하였다. 검증자료에 모형을 적용하여 임의시간대에서 사고 발생운전자 및 사고 미발생 운전자 집단의 분류 정확도를 분석하고, 임의 시간대 적용을 통하여 개발된 모형의 타당성 및 통계적 유의성을 검증하였다.

검증자료에 개발된 모형을 적용한 결과, 적중률(III)의 경우 사고 미발생 운전자의 분류 정확도는 기존 91.9%에서 검증결과 82.5%로, 적중률(IV)의 경

우 85.7%로 각각 낮아진 반면 사고발생 운전자의 분류 정확도는 검증결과 84.1%로 기존 76.7% 보다 높은 것으로 나타나 개발된 모형은 타당성을 가진다고 할 수 있다.

<Table 15> Verification of Accuracy Rate of Models (t3)

Classification model			Accuracy rate (III) (t3)				
Observation	Predictive value		Classification accuracy %	Observation	Predictive value		Classification accuracy %
	No occurrence	Occurrence			No occurrence	Occurrence	
No occurrence	79	7	91.9	No occurrence	52	11	82.5
Occurrence	10	33	76.7	Occurrence	10	53	84.1
Total			86.8	Total			83.3

<Table 16> Verification of Accuracy Rate of Models (t4)

Classification model			Accuracy rate (IV) (t4)				
Observation	Predictive value		Classification accuracy %	Observation	Predictive value		Classification accuracy %
	No occurrence	Occurrence			No occurrence	Occurrence	
No occurrence	79	7	91.9	No occurrence	54	9	85.7
Occurrence	10	33	76.7	Occurrence	10	53	84.1
Total			86.8	Total			84.9

<Table 17> Verification of Homogeneity of Predicted Classification Groups (t3~t4)

Verification variable	Predictive value		N	Average	Standard deviation	t	Significance probability
	No occurrence	Accuracy rate (III) (t3)					
XdYr	No occurrence	Accuracy rate (III) (t3)	62	0.03	0.016	-0.055	0.956
		Accuracy rate (IV) (t4)	64	0.03	0.016		
	Occurrence	Accuracy rate (III) (t3)	64	0.103	0.045	0.072	0.943
		Accuracy rate (IV) (t4)	62	0.103	0.045		

한편, 예측 분류된 집단의 시간대별 동질성에 대한 통계적 검증에서는 t-검정 결과 신뢰수준 95%($\alpha=0.05$)에서 시간대별로 차이는 유의하지 않은 것으로 나타났으며, 개발된 모형을 적용하였을 때 시간대별로 동일한 결과를 도출해 준다고 할 수 있다.

V. 결론 및 향후 연구과제

1. 결론

본 연구는 시내버스 운전자의 실제 운행기록 자료를 바탕으로 사고발생 운전자와 사고미발생 운전자를 분류할 수 있는 모형을 개발하는데 목적을 두었다. 본 연구를 위하여 2014년 1월~6월 중 사고발생 운전자 106명, 사고 미발생 운전자 149명 등 총 255명의 운행기록 자료를 수집·분석 하였으며, 판별분석(Discriminant Analysis) 및 로지스틱회귀분석(Logistic Regression Analysis)을 적용하여 운전자 사고발생 예측모형을 개발, 모형간 정확도를 비교하였다.

사고발생 유의변수 추출을 위해 임의 시간대별로 사고발생 운전자 집단 및 사고 미발생 운전자 집단의 운전행동 특성을 비교분석한 결과, 횡방향 가속도 등 6개 가속도 요인이 신뢰수준 95% 이상에서 그룹간 차이가 유의한 것으로 분석되었으며, 동 과정의 반복수행을 통하여 유의변수 선정의 타당성을 검증하였다. 이 결과를 토대로 횡방향 가속도 등 6개 가속도 변수를 사고발생 예측모형 개발을 위한 유의변수로 설정하였으며, 모형내 유의변수의 통계적 유의성 확보 및 모형에 대한 영향력이 낮은 변수의 순차적 제거를 통하여 사고운전자 분류 모형을 개발하고 개발된 각 모형의 분류정확도를 비교하였다.

모형개발 결과 총 30개의 모형이 개발되었으며, 이중 판별분석(Discriminant Analysis)에 의한 예측모형은 최대 62.8%의 비율로 사고발생 운전자 분류가 가능하였으며, 로지스틱회귀분석(Logistic Regression Analysis)에 의한 예측모형은 최대 76.7%의 비율로 사고발생 운전자 분류가 가능한 것으로 나타났다.

분석결과, 전반적으로 판별분석(Discriminant Analysis) 보다는 로지스틱회귀분석(Logistic Regression Analysis)을 적용 하였을때 사고발생 운전자의 분류율이 더 높게 나타났으며, 시내버스의 경우 6개 가속도 요인 중 진행방향에 대한 감속도(Xd) 및 우측방향 가속도(Yr)가 동시에 작용할 때 이 변수가 사고발생 운전자를 가장 잘 분류해 줄 수 있는 것으로 나타났다.

개발된 모형에 대한 타당성 검증을 위하여 임의 추출된 다른 시간대에 모형을 적용한 결과 76.7%의 적중률을 나타내었으며, 예측 분류된 집단에 대한 동질성 검정 결과에서는 신뢰수준 95%에서 집단간 차이는 통계적으로 유의하지 않은 것으로 나타나 개발모형의 타당성을 확보하였다. 또한, 개발된 모형의 추가 검증을 위하여 모형개발에 사용하지 않은 3월~6월 126명의 자료를 통하여 모형의 적합도를 분석한 결과, 사고발생 운전자 분류율은 기존 76.7% 보다 높은 84.1%의 적중률을 보이는 것으로 나타났으며, 예측 분류된 집단에 대한 시간대별 동질성 검정 결과 신뢰수준 95%에서 두 집단의 차이는 통계적으로 유의하지 않은 것으로 나타났다.

연구결과를 종합하면, 운전자의 운행기록 자료에서 수집되는 항목 중 진행방향에 대한 감속도(Xd) 및 우측방향 가속도(Yr)의 합성가속도 변수가 사고발생 운전자를 가장 잘 분류해 주는 것으로 나타났으며, 이는 향후 시내버스 운전자의 사고유발 가능성을 사전에 판단해 볼 수 있는 좋은 도구로 활용될 수 있을 것이다.

본 연구는 운전자의 실제 운행정보를 토대로 사고발생 가능성을 판단할 수 있는 방법론을 제시하는데 의의가 있다. 특히, 시내버스 사고의 높은 비중을 차지하고 있는 차내 안전사고의 감소를 위해 운전자 운전습관 교정의 전략적 근거를 제시하고, 급감속 및 급우회전 등의 운전행동 감소를 위한 운전자별 맞춤형 안전교육 시행의 이론적 틀을 마련하는 한편, 향후 신규 운전자 채용시 운행기록 분석을 통한 운전자 선별의 근거로 활용할 수 있을 것으로 기대된다.

2. 향후 연구과제

본 연구는 개인별 운행기록자료 수집의 한계로 개인별 운전행동을 정의하지 못하고 집단특성에 대해서만 비교되었으며, 운전자의 운전행동에 영향을 미칠 수 있는 도로 운행구간, 운행노선의 특성 등 외부요인에 대한 특성을 반영하지 못한 한계가 있다.

이에 개별 운전자의 운전행동에 대한 세부적인 규명을 통해 운전행동을 유형화하고, 노선별 운행 특성 및 구간별 기하구조 등 구체적인 도로여건을 수집하기 위해 교통사고 발생지점의 좌표를 이용하여 전자지도와 연계한 실증적인 연구와 자료구축을 통해 세부 모형을 추가로 개발할 필요가 있다. 또한, 본 연구에서는 운행기록 정보를 통하여 사고발생에 관계된 요인 규명에 관점을 두었으나, 과속, 급가속, 급출발 및 급감속 등 개별 위험운전행동과 교통사고와의 관계를 규명하는 연구 또한 본 연구의 연장선상에서 추가로 진행되어야 할 필요가 있다고 판단된다.

REFERENCES

[1] Bagadadi O. and Várhelyi A.(2011), "Jerky driving - An indicator of accident proneness," *Accident Analysis and Prevention*, vol. 43, pp. 1359-1363.

[2] Cho J. H.(2008), "Development of a Real-time Safe Driving Management System," University of Kookmin, Ph. D. Dissertation.

[3] Cho J. H. and Lee W. S.(2007), "Development of a Safe Driving Management System," *Transactions*

of Korea Society of Automotive Engineers, vol. 15, no. 1, pp.71-77.

[4] Hwang S. Ch.(2000), "A Study on the Causes of Traffic Casualties and Preventive Countermeasures," Dongguk University, Master's degree thesis.

[5] Lee H. S. and An B. J.(2006), "A Factor Analysis of Traffic Accidents Through Traffic Safety Diagnosis Results," *Journal of the Korean Society of Safety*, vol. 21, no. 2, pp.128-137.

[6] Oh J. S.(2011), "Discrimination for Reckless Drivers through Personality and Attitude Measurement : Developing a Psychological Test and Its Criteria for Driver Discrimination," Chungbuk National University, Ph. D. Dissertation.

[7] Oh J. T. and Lee S. Y.(2009), "A Study of the Weight value to Risky Driving Type," *International Journal of Highway Engineering*, vol. 11, no. 1, pp.105-115.

[8] Park J. H. and Lee S. Ch.(1997), "Analysis of Traffic Accident Proneness by Drivers' Variables," *Korean Journal of Culture and Social Issues*, vol. 3, no. 1, pp.143-157.

[9] af Wählberg. AE(2004), "The stability of driver acceleration behavior, and a replication of its relation to bus accidents," *Accident Analysis and Prevention*, vol. 36, pp. 83-92.

[10] Zegeer C., Huang H., Hummer J., Stutts J. and Rodgman E.(1993), "Characteristics and Solutions Related to Bus Transit Accidents," Federal Transit Administrations.

저자소개



김 정 열 (Kim, Jung Yeul)

1998. 2 : 안동대학교 물리학과 졸업

2004. 2 : 명지대학교 교통관광대학원 교통공학과 공학석사

2010. 2 : 명지대학교 대학원 교통공학과 박사수료

e-mail : dufgo@ts2020.kr



금 기 정 (Kum, Ki Jung)

1986. 2 : 명지대학교 토목공학과 공학사

1990. 3 : 동경도립대학교 대학원 토목공학과 공학석사

1993. 3 : 동경도립대학교 대학원 토목공학과 공학박사

1994 ~ 현재 : 명지대학교 공과대학 교통공학과 교수

e-mail : kjkum@mju.ac.kr