

로지스틱회귀분석 모델을 활용한 도시철도 사상사고 사고예측모형 개발에 대한 연구

Study on Accident Prediction Models in Urban Railway Casualty Accidents Using Logistic Regression Analysis Model

진수봉 · 이종우*

Soo-Bong. Jin · Jong-Woo. Lee

Abstract This study is a railway accident investigation statistic study with the purpose of prediction and classification of accident severity. Linear regression models have some difficulties in classifying accident severity, but a logistic regression model can be used to overcome the weaknesses of linear regression models. The logistic regression model is applied to escalator (E/S) accidents in all stations on 5~8 lines of the Seoul Metro, using data mining techniques such as logistic regression analysis. The forecasting variables of E/S accidents in urban railway stations are considered, such as passenger age, drinking, overall situation, behavior, and handrail grip. In the overall accuracy analysis, the logistic regression accuracy is explained 76.7%. According to the results of this analysis, it has been confirmed that the accuracy and the level of significance of the logistic regression analysis make it a useful data mining technique to establish an accident severity prediction model for urban railway casualty accidents.

Keywords : Accident prediction, Railway casualty accident, Logistic regression, Accident investigation, Falling accident

초 록 본 연구는 사고심각도 분류 및 예측을 위한 철도사고조사 통계기법에 관한 연구이다. 그동안의 선형 회귀분석은 사고 심각도 분석에 어려움이 있었으나 로지스틱회귀분석은 이를 보완할 수 있었다. 데이터마이닝 기법인 로지스틱회귀분석을 활용, 서울지하철(5~8호선) 역사 내 전도사고 중 에스컬레이터 전도사고 발생에 영향을 주는 사고예측 모형 변수는 사고자 연령, 음주여부, 사고 당시상황 및 행동, 핸드레일 잡음 여부였다. 분석의 정확도는 76.7%로 설명되었고 분석방법 결과에 따르면 정확도와 유의수준 측에서 로지스틱회귀분석 방법이 도시철도 사상사고 예측모형을 개발하는데 유용한 데이터마이닝 기법으로 판단된다.

주요어 : 사고예측, 철도사상사고, 로지스틱회귀분석, 철도사고조사, 전도사고

1. 연구배경 및 목적

서울시 도시철도 수송 분담률은 2013년 기준 타 교통수단 대비 약38.8%대로 가장 높은 비율을 차지하고 있고[1], 최근 서울메트로와 서울도시철도공사가 통합된 서울교통공사 일일 이용객은 약 800만명을 넘고 있으며 교통 정체 및 혼잡이 심해질수록 도시철도 이용수요는 증가될 전망이다. 도시철도를 이용하려면 역사 시설물을 가장 먼저 접하게 되는데 이동 과정에서 승객은 계단, 엘리베이터, 에스컬레이터 등 편의시설과 승강장안전문과 같은 안전시설에 의해 잠재적 철도사상사고에 노출된다.

Table 1은 2011년~2015년까지 서울지하철내 철도사상사고 중 철도사고 발생 1건당 인명피해(명) 정도를 나타내는데 여객사고가 직원에 비해서는 최대 39.5 배가 높았다. 대부분 사고는 편의시설 이용 시 발생하였고, 특히 노인승객 사고가 대부분(69%)이었으며, 고령화에 따른 노인 승객 증가로 사상 사고건수는 증가 추세이다. 운영기관에서는 철도사고조사 및 분석을 통한 사고감소에 노력하고 있으나 사고시간, 장소, 피해자 신상, 피해내용, 회수 등 기초통계에 의존한 원인분석이 대부분이어서, 고차원 통계분석을 통해 사고 원인별 상관관계분석, 심각도, 사고모형 예측 등 보다 신뢰성 높고 정교한 철도사고조사 분석 기법이 요구되고 있다.

따라서 Fig. 1과 같이, 철도사상사고 중 최근 5년간 사고 발생수와 사고자 수가 가장 많았던 에스컬레이터 전도사고(사고:399건, 사고자:410명)를 표본(Sample)으로 사고 유발요인에 대한 변수(Variable)를 도출하여 변수 간 사고에 미치는 특성분석과 사

고요인별 상관관계를 데이터마이닝(Data mining) 기법 중 로지스틱 회귀분석 모델을 통해 사고예측 모형을 개발함으로써, 철도 이용자 사고 예방 및 효과적인 안전대책에 기여하고자 한다.

Table 1. Accident risk by subject and line[2].

Year	Category	total	sub total	Subject			Risk by line			
				staff	public	passenger	5	6	7	8
average	Traffic accident	3.22	1.21	0.02	0.40	0.79	0.83	0.10	0.25	0.03
	Casualty accident		2.00	0.14	0.51	1.35	0.64	0.44	0.83	0.09
2015	Traffic accident	1.98	0.84	0.11	0.00	0.73	0.36	0.18	0.24	0.06
	Casualty accident		1.15	0.42	0.30	0.43	0.43	0.21	0.44	0.06
2014	Traffic accident	2.99	0.44	0.00	0.00	0.44	0.20	0.05	0.16	0.02
	Casualty accident		2.56	0.00	0.30	2.26	0.59	0.83	1.10	0.04
2013	Traffic accident	2.37	0.92	0.00	0.00	0.92	0.47	0.11	0.29	0.05
	Casualty accident		1.45	0.00	0.53	0.92	0.43	0.22	0.75	0.05
2012	Traffic accident	5.25	1.78	0.00	1.00	0.78	1.38	0.08	0.29	0.03
	Casualty accident		3.47	0.01	1.02	2.44	1.07	0.78	1.50	0.13
2011	Traffic accident	3.48	2.10	0.00	1.00	1.10	1.73	0.10	0.26	0.02
	Casualty accident		1.38	0.30	0.37	0.71	0.68	0.18	0.34	0.19

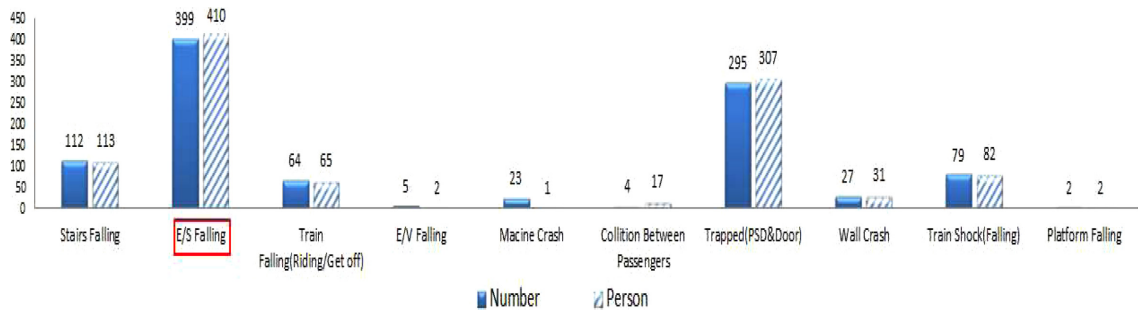


Fig. 1. Number of casualty accidents for pedestrian facilities.

본 연구에서는 서울도시철도(지하철) 역사 내 대합실, 승강장 등 철도시설 중 에스컬레이터 이용 여객의 전도사고(Falling Accident)에 영향을 주는 주요 사고유발 요인으로 한정하였고 크게 인적요인(사고자특성), 역사 환경 및 기하구조 요인, 사고자 행동 요인으로 구분하여 사고유발 요인을 독립변수로, 사고결과인 심각도 수준을 종속변수로 구분하여 각 독립변수 간 요인들이 종속변수인 사고 심각도에 어떤 인과관계 및 영향을 주는지 분석함으로써 사고감소 및 예방을 위한 효과적인 대책 마련을 제시하는데 연구목적이 있다.

에스컬레이터 전도사고의 사고예측모형을 개발하기 위해 Table 2과 같이 2014~2016년 3년간 서울교통공사 5~8호선 157개 역사내 에스컬레이터 전도사고 중 자체 사고조사 데이터를 대상으로, 이중 사고 원인이 불명, 기타 등 분석이 명확하지 않은 자료는 제거하였고 분석 가능한 사례만 표본추출(sampling)하여 전도사고와 관련이 있다고 판단한 14개 독립변수 항목을 선별하여 모형추적에 적합한 자료를 구축하였다.

Table 2. Data analysis resources of casualty accidents

Category	Contents
Period of data	2014 ~ 2016 year (E/S accident : 262 case)
Data source	Escalator safety accidents sampling data (Line 5~8 Seoul Metro 157 station internal resource)
Variable	Dependent variable (D.V) : 2 EA Independent variable (I.V) : 14 EA

2. 철도사고 예측모델 선행연구

위험도평가모델을 이용한 연구사례로 박찬우(2009)는 철도사상사고 시나리오 모형 구성 및 위험사건별 등가 사망지수로 환산된 정량적인 위험도 평가모델을 제시하였다.[3] 기초통계 및 설문조사 연구사례로 이종필(2009)은 지하철 이용승객 안전사고 감소방안에 대한 연구를 통해 5년간 발생한 지하철 이용 승객 사고분석과 승객들을 대상으로 안전 개선방안에 대한 설문조사를 실시하여 문제점에 대해 개선안을 마련함으로써 안전관리 방안을 연구하였다[4].

철도사상사고의 경향 및 대책 연구사례로 오재경(2014)은 서울메트로 철도안전사상사고의 연도별, 유형, 연령별, 성별, 이용승객별, 사고종류별 등 사고 발생 원인을 분석하고 이를 기반으로 사고를 예방할 수 있는 대안으로써 철도안전교육프로그램 및 철도안전체험센터 구축 필요성을 제시하였다[5]. 철도시스템 특성과 사고관계모델에 대한 분석연구사례로 강승규(1995)는 철도 건널목 사고예측모형 개발에 관한 연구를 통해서 안내원 유무가 사고 확률을 결정하는 가장 큰 요인으로 분석하였고, 결과로 열차속도, 철도와 도로 교통량, 교차각도, 선로수, 선로와 접근도로의 종단구배차 등 사고확률 결정 요인에 대해 연구하였다[6]. 도로교통 사상사고 관련 선형, 비선형 회귀분석방법 연구사례로 박주환(2012)은 교통사고 발생현황을 토대로 터널구간과 본선구간의 상호비교를 통한 사고원인별, 유형별 특성을 다중선형 회귀분석을 이용하여 터널내 교통사고 예측 모형을 개발하였고 [7], 박효신(2007)은 음이향 분포를 이용한 고속도로 인터체인지 연결로에서의 교통사고 예측모형을 개발하여 연결로 형식별 사고위험도를 분석하였다[8].

또한 C. V. LZeger(1986)는 교통사고 원인별 독립변수들에 대한 단계별 선형회귀분석을 통해 변수간 상관성을 조사하였으며, 비선형 회귀분석을 통해 지방 2차선 도로에 대한 사고예측모형을 개발하여 도로이 확장될 경우 사고 감소율을 도출하였다 [9]. Table 3과 같이 최근 각 교통수단별로 각종 통계기법을 통해 데이터 간 유의한 해석을 도출하려는 연구가 다양하게 시도되었으나 그동안의 연구사례는 사고가 언제, 어디서, 어떻게, 어떤 사유로 발생했는지에 대한 기초(빈도)분석 및 이용자 대상 설문조사에 의한 연구시행과 관찰된 연속형 변수로 이루어진 2~3개의 독립변수로 유의율을 판단하는 선형회귀분석 기법 등을 활용한 연구가 대다수였다. 기존의 다중회귀분석은 선형회귀분석의 일종으로 종속변수가 양적인(quantitative) 척도 (즉, 간격척도나 비율척도)로 측정된 경우의 분석방법이었다[10].

에스컬레이터 전도사고는 사망, 중상, 출혈 등 심각사고와 통증호소 후 귀가 등 경미사고로 분류할 수 있고, 이 분류된 사고 결과에 대한 원인이 되는 독립변수는 대부분 양적인 척도가 아닌 질적인(qualitative) 범주형 척도로서, 이를 크게 「원인-결과」 이분형 변수(binary variable)로 나타낼 수 있다. 따라서 원인에 해당되는 독립변수와 결과에 해당되는 종속변수 간 관계를 나타내는 이분형 변수(binary variable)형태인 로지스틱회귀분석을 사용하여 사고를 분석할 필요가 있다. 독립변수를 이용한 종속변수의 예측은 선형회귀분석을 통해서도 가능하지만 종속변수가 이분형 변수인 경우에는 선형회귀분석을 이용하여 독립변수와 종속변수간의 관계를 분석할 수 없다. 이유는 선형회귀분석을 적용하기 위해서는 종속변수가 정규분포, 등분산성, 독립성을 따르는 연속형 변수이어야 하기 때문이다[11].

Table 3. Casualty accidents and prediction models related research

Category	Contents	Author
Automobile	Analysis of traffic accident factor using decision tree model	M.Y.Lee (2013)
	Development an accidents forecasting models in freeway using multiple linear regression analysis	J.H.Park (2012)
	Severity analysis of the pedestrian crash patterns based on the ordered logit model	J.S.Choi (2009)
	Development of accident prediction models for freeway interchange ramps	H.S.Park (2007)
Railway	Developing an accidents prediction model for railroad - highway grade crossings	S.K.Kang (1995)
	A Factor analysis of urban railway casualty accidents and establishment of preventive response systems	J.K.Oh (2014)
	A Study on the passenger's accidents in the subway	J.P.Lee (2009)
	Development of risk assessment model for railway casualty accidents	C.W.Park (2009)

따라서 도시철도(지하철) 역사 내 발생빈도가 가장 많은 에스컬레이터 전도사고를 대상으로 로지스틱회귀분석을 통한 사고모형 개발과 정확도를 분석하여, 기존 선행 연구 방법보다 정확도가 높고 고차원적 데이터 분석 방법을 연구하는 것은 철도사고 조사 분야에서 매우 큰 의미가 있다.

3. 에스컬레이터 전도사고의 로지스틱회귀분석 모델

3.1 로지스틱회귀분석 모델

로지스틱회귀분석(Logistic Regression Analysis)은 특정 사고가 발생할지 여부를 직접 예측하는 것이 아니라 그 사고가 발생할 확률을 추정하는 분석 방법의 일종이다. 종속변수의 예측 값은 Fig. 2와 같이 항상 0과 1사이 확률 값을 갖게 되며 값이 0.5보다 크면 심각 사고발생 확률이 높고 0.5보다 작으면 심각 사고발생 확률이 낮다고 예측할 수 있다. 선형 회귀분석에서는 독립변수와 종속변수간의 선형관계를 가정하지만 로지스틱회귀분석에서는 사고 원인이 되는 독립변수와 사고결과인 종속변수 간 비선형인 S곡선의 형태를 갖는다.

로지스틱 회귀분석은 독립변수와 이분형 종속변수 간의 관계를 분석하는데 있어서 다른 회귀분석 방법에 비해 매우 유연하게 사용할 수 있어 본 연구에서는 로지스틱회귀분석(로짓 분석:logit analysis이라고도 함)을 사용하였다.

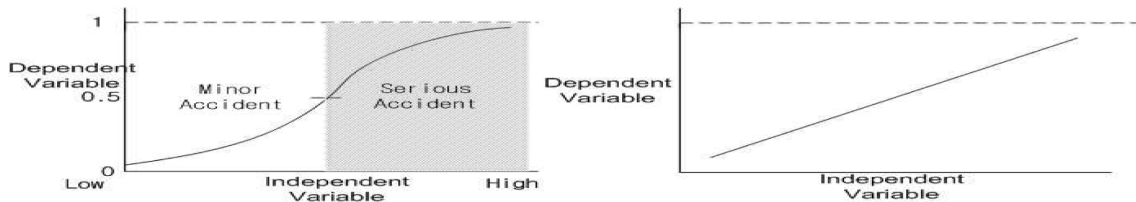


Fig. 2. Plot of logistic & simple/multiple regression [12]

로지스틱회귀분석에서는 사고가 일어날 확률을 추정한다. 만약 독립변수가 하나일 경우 로지스틱 회귀모델 (Logistic regression model)은 ‘식(1)’과 같다.

$$Prob(\text{심각사고발생}) = \frac{1}{1 + e^{-(b_0 + b_1x)}} \quad (1)$$

여기서, Prob(사고발생:종속변수) : 사고가 발생할 예측 확률, b_0 와 b_1 : 로지스틱회귀계수로 (+) 또는 (-)값을 취함
 x : 독립변수, e 는 자연로그의 밑수로서 약 2.718

본 논문에서는 독립변수가 하나 이상이므로 로지스틱회귀모델은 ‘식(2)’, ‘식(3)’ 과 같이 표현될 수 있다.

$$Prob(\text{심각사고발생}) = \frac{1}{1 + e^{-z}} \quad (2)$$

여기서, $z(\text{독립변수의 선형결과 값}) = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$
 p : 독립변수(역사 내 전도사고 심각도에 영향을 미치는 변수)의 개수

$$e^{-z} = \frac{1 - Prob(\text{심각사고발생})}{Prob(\text{심각사고발생})} = \frac{Prob(\text{경미사고발생})}{Prob(\text{심각사고발생})} \quad (3)$$

$$[z] = \ln \frac{Prob(\text{심각사고발생})}{Prob(\text{경미사고발생})} \quad (4)$$

또한 ‘식(4)’와 같이 “심각사고가 발생할 확률 대 경미사고 발생확률 비율의 로그값”으로 표현되며 이 식에서 $\left(\frac{Prob(\text{심각사고발생})}{Prob(\text{경미사고발생})}\right)$ 을 오즈(odds)라 하고 이 로그 값을 취한 값을 ‘식(5)’와 같이 표현할 수 있다.

$$[z] = \ln \frac{Prob(\text{심각사고발생})}{Prob(\text{경미사고발생})} = \ln Odds = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (5)$$

따라서 ‘식(5)’는 본 연구의 전도사고 예측 모형식으로 활용 할 수 있으며 회귀계수가 0보다 크면 오즈비는 1보다 크고 회귀 계수가 0보다 작으면 오즈비는 1보다 작게 된다. 회귀계수가 0이면 오즈비는 1이 된다. 특정 독립변수에 대한 오즈비가 1보다 크다는 의미는 그 독립변수를 한 단위증가(본 논문에서는 0→1)시키면 오즈, 즉 심각한 사고 발생 비율이 증가한다는 뜻이며 이는 특정 독립변수에 대한 오즈비 값이 큰 값 일수록 심각 사고가 발생할 가능성이 높아진다는 의미이다. 반대로 특정 독립변수에 대한 오즈비가 1보다 작다는 의미는 그 독립변수를 한 단위 증가시키면 오즈비 즉 심각한 사고가 발생할 가능성이 낮아진다는 의미이다.

3.2 철도사고 분류 및 가설

철도사고는 철도교통사고와 철도안전사고로 나누며 이중 철도사상 사고는 Table 4와 같이 분류한다. 본 논문에서의 철도사상 사고는 「여객이 에스컬레이터 이용 중 발생한 전도사고」로 한정한다.

Table 4. Railway accident classification

Railway accident	Railway traffic accident	Railway accident	
		Crossing accidents	
		Railway traffic casualty accident	
	Railway safety accident	Railway fire accident	
		Railway (safety) casualty accident	passenger *
			public
			staff
Railway facility breakage accident			
Other Railway safety accident			

본 연구에서는 Table 5와 같이 전도사고 발생에 영향을 미칠 것으로 판단되는 변수를 독립변수로 선정하였고, 이를 기준으로 사고 발생 가설을 크게 3가지로 수립하였다. 가설I은 「사고자 인적요소로 인해 사고 심각도에 영향을 준다.」라는 가설이다. 즉, 사고자 성별, 연령, 음주여부, 개인의 장애유무(장애우) 등에 따라 사고 발생 심각도 차이를 검증하고자 하였다. 가설II는 「이용 환경 및 시설물 구조적 요소로 인해 사고 심각도에 영향을 준다.」라는 가설이다. 즉, 사고 발생 시간, 날씨, 역사 구조적 특징 등에 따라 심각도 차이를 검증하고자 하였다. 가설III은 사고당시 사고자가 어떤 행동을 취하고 있었는지 즉, 「사고당시 사고자 행동이 사고 심각도에 영향을 준다.」라는 가설이다.

3.3 심각도 모형 개발 및 적합성 분석

본 연구에서는 에스컬레이터 전도사고와 관련 있다고 예측된 다양한 독립변수를 사용하여 전도사고 예측모형을 개발하는 것으로서 여러 독립변수를 사용하고자 로지스틱회귀분석 후 유의성 있는 변수들을 추출하였다.

각 철도사고조사 데이터는 회귀분석 프로그램인 SPSS 21.0 Software를 이용하여 분석하였으며, 기초통계를 통해 조사대상의 사고현황을 파악하였고 관련 독립변수 14개를 모두 넣고 회귀분석을 한 후 유의하지 않다고 나온 변수는 제거하면서 최종 모형을 구축하였다. 사고 심각도는 3년간 사망사고가 2건으로 매우 적어 중상, 출혈 등의 사례를 심각사고로, 그 외 통증 호소 후 귀가 등의 사례는 경미사고로 구분하였다.

총 14개 독립변수 중 최종적으로 유의수준이 0.05보다 작아 사고예측 모델 구축에 유의하다고 분석된 5개 독립변수를 이용하여 에스컬레이터 전도사고 심각도 모형을 구축한 결과는 Table 6과 같았다.

로지스틱 회귀모형의 가설검정은 전체 로지스틱 회귀모델의 유의성 검정, 개별 로지스틱 회귀계수에 대한 유의성 검정의 두 가지 측면에서 수행한다. 전체 로지스틱 회귀모델 유의성검정은 우도비(likelihood ratio)라는 통계량으로 수행한다. 또한 최대 우도법(maximum-likelihood method)라는 방법을 이용하여 모수(로지스틱 회귀계수)를 추정한다. 여기서 모수 추정치가 주어졌을 때 관측 값이 나타날 확률을 우도(likelihood)라고 하고 이 값의 로그 값에 -2배를 한 지표(-2LL, -2log-likelihood)가 추정된 모델이 데이터를 얼마나 적합 시키는지 척도로 사용한다.

첫 번째로 우도비 검정으로 우도비는 모델에 변수가 추가되거나 제거될 때의 모델 적합도의 변화량을 나타내며 -2LL값의 차이로 표현한다. Table 7에서는 카이 제곱 값이 최종적인 검정통계량이 되며 값은 119.402, 자유도는 5이고, 유의수준이 0.05값

Table 5. Variables used in Accident prediction models and frequency analysis

Category	Variable	Data coding	Basic statistics(n)	Nominal variable		
I.V	Human factor	Gender	female : 0	112	•	
			male: 1	150		
		Age(year)	~ 65 : 0	119	•	
			over 65: 1	143		
		Drinking	non-drinking : 0	160	•	
			drinking : 1	102		
	Disabled person	no : 0	250	•		
		yes : 1	12			
	environment & geometric factor	Accident time	non-rush hour : 0	179	•	
			rush hour (07~09, 18~20) : 1	83		
		Weather	rain or snow : 0	143	•	
			clear : 1	119		
		Transfer station (Confusion Station)	no : 0	132	•	
			yes : 1	130		
		Escalator type	separate : 0	196	•	
			integrated : 1	66		
		Curved platform	straight : 0	58	•	
			curve : 1	204		
		behavior factor	Situation (direction of movement)	downward : 0	193	•
				upward : 1	69	
	Handrail grip		yes : 0	148	•	
no : 1			114			
Behavior	stand : 0		185	•		
	walk, run : 1		77			
Carelessness	other : 0		28	•		
	self : 1		234			
Belongings	no : 0		120	•		
	yes : 1		142			
D.V Accidet severity	Minor	no external injury : 0	127	•		
	Serious	death, serious injury, bleed : 1	135			

Table 6. Results of logistic regression analysis

	B	S.E.	Wals	df	sig.	Exp(B)*	95% Confidence interval of EXP(B)		
							Lower	Upper	
1Step ^a	Age(x1)	.903	.340	7.072	1	.008	2.467	1.268	4.799
	Situation(x2)	-1.490	.405	13.557	1	.000	.225	.102	.498
	Drinking(x3)	2.211	.397	31.042	1	.000	9.126	4.192	19.865
	Behavior(x4)	.806	.376	4.581	1	.032	2.238	1.070	4.679
	Handrail grip(x5)	2.192	.340	41.484	1	.000	8.958	4.597	17.456
Constants	-1.982	.387	26.188	1	.000	.138			

B: Regression coefficient, S.E. : Standard error, Wald : (B/S.E)², df : Degree of freedom, Sig. : Significance level, Exp(B) : Odds ratio

Table 7. Verification of likelihood

		χ^2	df	sig.
1 Step	Step	119.402	5	.000
	Block	119.402	5	.000
	Model	119.402	5	.000

보다 작으므로 통계분석 결과는 매우 유의한 것으로 나타났다.

두 번째로 사고발생에 대한 관측빈도와 로지스틱회귀모델의 예측빈도간 적합도를 평가하는 Hosmer와 Lemeshow 검정의 p값이 Table 8과 같이 0.589로 유의수준 0.05보다 크므로 본 연구 모형은 통계적으로 적합으로 분석되었다.

Table 8. Verification of Hosmer and Lemeshow zz

Step	χ^2	df	sig.
1	6.524	8	.589

세 번째로 선형회귀분석의 R제곱에 대응하는 종속변수의 분산비율을 나타내는 유사결정계수 값이 Table 9와 같이 0.366, 0.488로 독립변수들로 이루어진 회귀식이 종속변수를 예측 할 수 있는 설명력이 48.8%로 나타났다.

Table 9. Verification of Cox @ Snell R² & Nagelkerke R²

Step	-2Log Likelihood	Cox @ Snell R ²	Nagelkerke R ²
1	243.563 ^a	.366	.488

Table 10과 같이 최종 로지스틱 회귀분석의 정확도 분석 결과, 유의수준 5%에서 총 14개 변수 중 5개 변수(Age, Situation, Drinking, Behavior, handrail grip)가 유의수준을 보였으며 분류 정확성은 76.7%로 높게 분석되었다.

Table 10. Data analysis resources for casualty accidents

Watched	Logistic regression analysis (n=262)		
	Minor	Serious	Accuracy(%)
Minor	99	28	78.0
Serious	33	102	75.6
Total(%)			76.7

4. 사고예측 모형 및 적합성 검증

3장에서 설명한 바와 같이 개발된 모형을 평가한 결과 최종 선정된 에스컬레이터 전도사고 심각도 예측모형 식은 Table 11과 같다. 「상황(situation)」에 대한 독립변수만 제외하고 모두 정(+)의 영향을 받고 있다. 즉 독립변수를 한 단위 증가(0→1)시킬 때 사고 심각도는 「상황(situation) 독립변수」를 제외하고 모두 증가되는 것으로 분석되었다.

Table 11. Accident predicting model equation

Accident prediction Model expression	$z = \ln \frac{\text{Prob}(\text{Serious accidents})}{\text{Prob}(\text{Minor accidents})} = \ln \text{Odds} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$ $= -1.982 + [2.192 \times \text{handrail grip (x5)}] + [2.211 \times \text{drinking(x3)}]$ $+ [0.806 \times \text{behavior(x4)}] + [0.903 \times \text{age(x1)}] - [1.490 \times \text{situation(x2)}]$
--------------------------------------	--

전도사고 예측모형의 해석은 Table 6에 있는 (Exp(B)), 즉 오즈비를 사용하여 해석했다. 모형 결과를 토대로 사고모델설정 즉, 65세 이상 노년층이 핸드레일을 잡지 않고 음주상태로 상행 에스컬레이터에서 걸거나 뛰는 행동으로 이용하다 전도사고 발생

시 심각사고로 발생할 사건의 독립변수의 선형 결정값 Z(a)와 다른 변수들은 변화가 없는 상태에서 「핸드레일 잡음」 변수만 변화, 즉 핸드레일을 잡고 이용했다는 조건만 변경했을 때 독립변수의 선형 결정값 Z(b)은 Table 12와 같이 표현할 수 있다.

Table 12. Data analysis resources for casualty accidents

Z(a) Accident prediction model (I)	Z(b) Accident prediction model (II)
$Z = -1.982 + 2.192 + 2.211 + 0.806 + 0.903 - 1.490 = 2.640$	$Z = -1.982 + 2.192 \times 0 + 2.211 + 0.806 + 0.903 - 1.490 = 0.448$
$Prob = \frac{1}{1 + e^{-2.640}} = 0.9334 = 93.34\%$	$Prob = \frac{1}{1 + e^{-0.448}} = 0.6101 = 61.01\%$
$z = \ln \frac{0.9334}{0.0666} = \ln 14.015 = 2.640$	$z = \ln \frac{0.6101}{0.3899} = \ln 1.565 = 0.448$

여기서 「핸드레일 잡지않음→잡음」으로 독립변수가 변할 때 Z값이 2.640에서 0.448로 변한 차이 B값은 2.192로 Table 6의 핸드레일(x5) 회귀계수와 일치한다. 또한 오즈비 변화율 Exp(B)은 핸드레일을 잡지 않고 타는 것이 잡고 타는 것에 비해 사고 심각도는 $z = \ln \frac{14.015}{1.565} = 8.958$ 배 높으며 역시 Table 6의 Exp(B)*=8.958값과 일치한다.

5. 결 론

본 연구에서는 데이터마이닝 기법 중 로지스틱회귀분석 모델을 통해 철도사상사고 중 에스컬레이터 전도사고에 영향을 주는 변수를 선택하여 사고예측 모형을 제시하였고, 통계 검증을 수행하여 신뢰성 및 정확성을 검증하였다.

이는 철도사고조사 기초 데이터를 이용하여 사고 특성을 분석하고 통계적 유의성을 확보한 변수를 활용하여 사고예측 모형 개발을 시도하였다는 점에서 큰 의의가 있다. 본 연구 분석을 통해 다음과 같이 전도사고(에스컬레이터)의 주요 특징을 도출할 수 있었다.

첫째, 65세 이상 노년층이 하행 방향 에스컬레이터에 음주 상태에서 핸드레일을 잡지 않고 걷거나 뛰는 행동으로 이용하는 경우 등 인적, 사고자 행동요인이 전도사고 심각도에 가장 높은 영향을 주는 것으로 나타났다. 이는 고령화 인구의 도시철도(지하철) 이용률 증가에 따라 편의시설 이용 시 본인 부주의로 인한 안전사고가 늘다는 것을 의미한다. 신체적 노화로 인한 활동 능력 및 반사 신경 둔화, 인지능력이 떨어지는 상태에서 음주상태 후 에스컬레이터 스텝(step)의 헛디딤이나 스텝의 빗살 줄무늬에 대한 시각적 현기증으로 인한 넘어짐, 구름사고 등 위험에 노출되기 쉽고 이로 인해 부상 심각도가 상대적으로 높아지는 것으로 분석되었다.

둘째, 시간, 날씨, 환승역, 승강장형태, 곡선승강장, 이용객의 소지품 소유 유무 등 환경 및 기하학 구조 등 기타 요인변수는 사고 심각도에 크게 유의하지 않은 것으로 분석되었다. 이는 철도운영자 측에서도 안전장치 확충 등 사고 방지를 위해 노력해 야겠지만 이용자 부주의와 비정상적 이용행태에 따라 발생하는 1차적인 안전사고 발생방지를 위해 무엇보다도 이용객들을 대상으로 각종 매체를 활용한 안전교육 등 사고예방을 위한 홍보 및 안내방송을 보다 강화할 필요가 있다고 판단된다. 본 연구의 목적은 사고예측 모형 개발이었으나, 향후 사고 심각도에 영향을 미칠 수 있는 밝혀지지 않은 독립변수들을 지속적으로 보완, 발굴하여 보다 완벽하고 체계적인 모형 구축을 위한 연구가 수행되길 기대한다. 데이터마이닝 기법은 수집된 데이터의 크기, 형태, 가설설정 등 여러 조건들에 따라 데이터 분석방법이 달라지기 때문에 해당 사례에 적합한 기법을 선택해야 한다. 따라서 향후 본 연구 방법을 기초로 로지스틱회귀분석 외 의사결정나무 모델 및 신경망 모형 분석, 구조방정식 모델 등 추가 데이터마이닝 분석방법을 활용하여 승강장에서 발생하는 열차 출입문 및 승강장 안전문 기입 사고, 계단전도사고 등 사고예측 모형 분석은 물론, 더 나아가 시스템 고장예측, 시스템 잔존 수명평가 예측 등 향후 유사 연구에 데이터마이닝 기법을 적용함으로써 좀 더 신뢰성 있고 유의성 높은 모형개발이 될 수 있으리라 기대한다.

References

- [1] www.seoul.go.kr (Accessed 29 June 2015)
- [2] T.H. Kim (2015) Accident risk by subject and line, SMRT, The report of risk analysis Seoul Metro 5-8 line.
- [3] C.W. Park, J.B. Wang, M.S. Kim, D.B. Choi et al. (2009) Development of risk assessment model for railway casualty accidents, *Journal of Korean Society For Railway*, 12(2), pp.190-198

- [4] J.P. Lee (2009) A Study on the passenger's accidents in the subway, MS Thesis, University of Seoul
- [5] S.G. Kim, I.H. Park, J.K. Oh, Y.K. Kim et al. (2014) A Factor analysis of urban railway casualty accidents and establishment of preventive response systems, *Journal of Korean Society Of Civil Engineers*, 34(3), pp.1017-1022
- [6] S.K. Kang (1995) Developing an accidents prediction model for railroad - highway grade crossings, *Journal of Korea Transportation Research Society*, 13(2), pp.43-58
- [7] J.H. Park, S.K. Kim (2012) Development an accidents forecasting models in freeway using multiple linear regression analysis, *The Journal of The Korea Institute of Intelligent Transport Systems*, 11(6), pp.145-154
- [8] H.S. Park, B.S. Son, H.J. Kim (2007) Development of accident prediction models for freeway interchange ramps, *Journal of the Korean Society of Road Engineers*, 25(3), pp.123-135
- [9] C. V. Zegeer, J. Hummer, D. Reinfurt, L. Herf and W. Hunter (1986) Safety effects of cross-section design for two-lane roads, *Federal Highway Administration*, FHWA-RD-87-008
- [10] H.S. Lee, J.H. Lym (2012) *SPSS 18.0 Manual*, Jyp Hyun Jae, Seoul, pp.342-347
- [11] K.Y. Kim, M.S. Jeon, H.C. Kang, S.K. Lee (2009) *Regression analysis by example*, Free Academy, Seoul, pp.334-357
- [12] H.Y. Lee, S.C. No (2009) *Advanced statistical analysis (Theory and practice)*, Bum Mun Sa, Seoul, pp.357-395

(Received 17 August 2017; Accepted 25 August 2017)

Soo-Bong Jin : sb19667@naver.com

Department of Railway Electrical and Signaling Engineering, Graduate School of Railway, Seoul National University of Science and Technology, 232 Gongreungro, Nowon-gu, Seoul, 139-743, Korea

Jong-Woo Lee : sanganlee@seoultech.ac.kr

Department of Railway Electrical and Signaling Engineering, Graduate School of Railway, Seoul National University of Science and Technology, 232 Gongreungro, Nowon-gu, Seoul, 139-743, Korea