

# 딥 러닝을 이용한 고속도로 교통사고 건수 예측모형 개발에 관한 연구

## A Study for Development of Expressway Traffic Accident Prediction Model Using Deep Learning

류종득\* · 박상민\*\* · 박성호\*\* · 권철우\*\*\* · 윤일수\*\*\*\*

\* 주저자 : 한국도로공사 수도권건설사업단 단장  
\*\* 공저자 : 아주대학교 건설교통공학과 박사과정  
\*\*\* 공저자 : 아주대학교 건설교통공학과 석사과정  
\*\*\*\* 교신저자 : 아주대학교 교통시스템공학과 부교수

Jong-Deug Rye\* · Sangmin Park\*\* · Sungho Park\*\* · Cheolwoo Kwon\*\*\* · Ilsoo Yun\*\*\*\*

\* Construction Office., Korea Expressway Corporation  
\*\* Dept. of Transportation Eng., Ajou University  
\*\*\* Dept. of Transportation Eng., Ajou University  
\*\*\*\* Dept. of Transportation Eng., Ajou University  
† Corresponding author : Ilsoo Yun, ilsooyun@ajou.ac.kr

Vol.17 No.4(2018)  
August, 2018  
pp.14~25

ISSN 1738-0774(Print)  
ISSN 2384-1729(On-line)  
<https://doi.org/10.12815/kits.2018.17.4.14>

Received 10 June 2018  
Revised 4 July 2018  
Accepted 8 July 2018

© 2018. The Korea Institute of  
Intelligent Transport Systems. All  
rights reserved.

### 요 약

최근 빅데이터 시대의 도래와 함께 교통사고와 관련된 요인을 설명하기 용이해졌다. 이에 따라 최신 분석 기법을 적용하여 교통사고 자료를 분석하고 시사점을 도출할 필요가 있다. 본 연구의 목적은 고속도로 교통사고 자료를 이용하여 고속도로의 주요 분석 단위인 콘존의 교통사고 건수를 예측하기 위하여 음이항 회귀모형과 딥 러닝을 이용한 기법을 적용하고 예측 성능을 비교하였다. 예측 성능 비교 결과, 딥 러닝 모형의 MOE들이 음이항 회귀모형에 비해 다소 우수한 것으로 나타났으나, MAD 기준으로 차이는 미미한 것으로 나타났다. 하지만 딥 러닝을 이용할 경우 다른 독립변수들을 추가하는 것이 용이하고, 모형의 구조 등을 변경할 경우 예측 신뢰도를 더욱 증가시킬 수 있을 것으로 판단된다.

핵심어 : 교통사고 건수 예측모형, 딥 러닝, 교통안전, 안전성능함수, 음이항 회귀모형

### ABSTRACT

In recent years, it has become technically easier to explain factors related with traffic accidents in the Big Data era. Therefore, it is necessary to apply the latest analysis techniques to analyze the traffic accident data and to seek for new findings. The purpose of this study is to compare the predictive performance of the negative binomial regression model and the deep learning method developed in this study to predict the frequency of traffic accidents in expressways. As a result, the MOEs of the deep learning model are somewhat superior to those of the negative binomial regression model in terms of prediction performance. However, using a deep learning model could increase the predictive reliability. However, it is easy to add other independent variables when using deep learning, and it can be expected to increase the predictive reliability even if the model structure is changed.

Key words : Traffic accident prediction model, Deep learning, Traffic safety, Safety performance function, Negative binomial regression model

## I. 서 론

### 1. 연구의 배경 및 목적

2016년 우리나라에서는 총 220,917건의 교통사고가 발생하였다. 이로 인한 사망자는 총 4,292명이고, 부상자는 331,720명으로 집계되었다(National Police Agency, 2017). 최근 각고의 노력을 통해 우리나라의 사망자 수는 줄어드는 추세에 있다. 도로부문에서는 자동차 등록대수가 2011년 1,844만대에서 2015년 2,099만대로 연평균 3.3% 증가하였음에도 불구하고, 교통사고 사망자 수가 연평균 3.9%로 감소하고 있으며, 자동차 1만 대당 사망자 수도 29.2% 감축되는 큰 성과를 거두고 있다. 하지만, 여전히 OECD 회원국에 비하여 높은 사망자 수 및 사망률을 보이고 있는 실정이다. 이에 정부는 국가 교통안전 향상을 위해 향후 5년간(2017년~2021년) 추진 예정인 「제8차 교통안전기본계획」을 통해 2021년까지 교통사고 사망자 수를 2,700명대(2,796명)로 줄여 나가겠다고 밝히고 있다. 이를 위해 “5030”으로 불리는 도심부 제한속도의 단계적 하향 등 다양한 사업을 추진 중에 있다.

이러한 도로부문에서 교통안전성을 높이기 위해서는 우선 각종 도로 상에서 발생하는 교통사고 건수에 대한 정확한 예측이 선행되어야 한다. 이러한 예측을 통해 선제적으로 위험 도로 구간을 선별하여 기하구조 등을 개선할 수 있으며, 도로를 위험도에 따라서 분류함으로써 체계적인 관리도 가능해진다. 최근 교통사고 자료에 대한 관심이 높아지고 있으며, 교통사고 자체의 정확도 및 정밀도를 높이려는 노력과 함께 교통사고 자료를 이용한 각종 교통안전성 분석이 활발하게 진행되고 있다. 특히 빅데이터 시대의 도래와 함께 교통사고와 관련된 요인을 설명할 수 있는 자료들의 종류와 양이 늘어나고 있어 최신 분석 기법을 이용하여 기술적으로 보다 정밀하게 교통사고 자료를 분석하고 시사점을 도출할 수 있게 되었다. 기존에는 교통사고 건수를 예측하는 모형 구축을 위해서 전통적인 포아송 회귀모형과 음이항 회귀모형 등을 사용하여 왔다. 하지만, 최근 딥 러닝 등 머신 러닝 기법을 이용한 접근 방법들이 주목을 받고 있으며, 우리나라에서도 이러한 적용이 활발하게 진행되고 있다.

이에 본 연구에서는 교통사고 자료 및 각종 기하구조 자료들이 상대적으로 관리 및 집계가 잘 되고 있는 고속도로를 대상으로 딥 러닝 기법을 이용하고 교통사고 건수를 예측하고자 한다. 참고로 국가기간교통망을 형성하고 있는 고속도로는 현재 28개 노선 3,989km에 이르고, 연간 154,033만대가 고속도로를 이용하고 있다. 2016년 한 해 동안 고속도로에서는 2,195건의 교통사고가 발생하였으며, 이로 인해 총 239명이 사망하고, 1,424명이 부상을 당하였다. 본 연구에서는 고속도로 기하구조 및 교통사고 자료를 콘존(congestion zone) 단위로 가공하여 고속도로 교통사고 건수 예측모형을 딥 러닝 바탕으로 개발하였다. 여기서 콘존이라 함은 한국도로공사가 고속도로 구간을 IC(interchange), JCT(junction), 그리고 TG(tollgate) 등 통행하는 차량수가 일정한 고속도로 구간으로 분류한 개념으로서 교통소통 및 안전과 관련된 자료들이 수집, 저장 및 제공되는 기본 단위 중에 하나이다(Ryu, 2018). 또한, 개발된 딥 러닝 기법 기반 고속도로 교통사고 건수 예측모형을 기존 전통적인 통계 기반 예측모형과 성능을 비교함으로써 딥 러닝을 이용한 고속도로 교통사고 건수 예측모형의 예측 성능을 검증하고자 한다.

### 2. 연구의 범위 및 방법

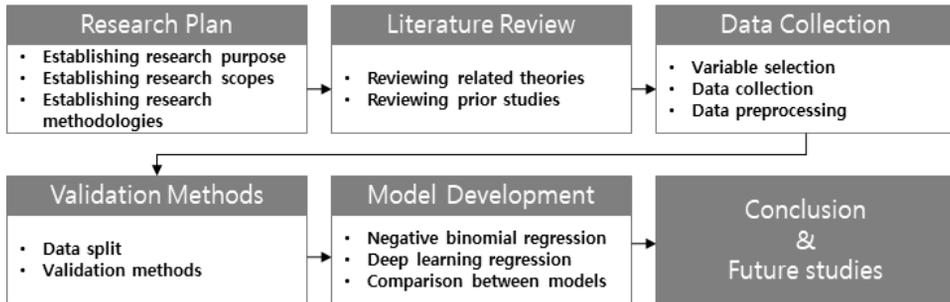
#### 1) 연구의 범위

본 연구의 공간적 범위는 자료 수집의 용이성을 고려하여 우리나라의 고속도로 중 한국도로공사가 관리

하고 있는 고속도로 및 민자 고속도로로 한정한다. 또한 시간적 범위는 2013~2015년까지 3년간의 고속도로 교통사고, AADT, 기하구조, 시설물 자료를 활용하고자 한다.

## 2) 연구의 방법

본 연구는 관련 이론 및 선행연구 고찰, 고속도로 교통사고 자료 수집, 모형 검증 방법 선정, 전통적 통계 모형을 이용한 고속도로 교통사고 건수 예측모형 개발, 딥 러닝을 이용한 고속도로 교통사고 건수 예측모형 개발, 모형 성능 비교, 결론 도출 순으로 수행하였다.



〈Fig. 1〉 Research process

## II. 관련 이론 및 연구 고찰

### 1. 관련 이론 고찰

#### 1) 교통사고 예측모형

교통사고 건수 예측을 위해서 다양한 형태의 모형들이 사용돼 왔다. 이 중 최근에 가장 활발하게 사용되는 교통사고 건수 예측모형은 안전성능함수(Safety Performance Function, SPF)라고 불리는 미국 도로안전편람(Highway Safety Manual, HSM)에서 제시하고 있는 모형이라고 판단된다. SPF는 일반적으로 이상적인 상태에서 해당 도로의 교통사고 발생빈도를 예측 또는 추정하는데 활용되며, 연평균일교통량(AADT)과 도로구간연장 등을 노출계수(exposure)로 하여 곡선반경 등 다양한 설계 요소를 독립변수로 반영하여 산출한다(Oh and Kang, 2017). 특히, SPF는 사용하는 독립변수의 종류에 따라 단순 안전성능함수(simple SPF)와 통합 안전성능함수(inclusive SPF)로 구분된다. 단순 안전성능함수는 독립변수로서 일평균교통량(ADT) 또는 연평균일교통량(AADT)를 사용하여 모형을 구축하며, 식(1)과 같은 형태를 가진다(Korea Expressway Corporation, 2014).

$$N_{SPF} = e^{\alpha} \times ADT^{\beta} \quad (1)$$

여기서,  $N_{SPF}$  = 예측사고건수 또는 예측사고율(건/km/년)

$ADT$  = 해당 구간의 일평균 교통량(대/일)

$\alpha, \beta$  = 회귀 계수

통합 안전성능함수는 교통량 및 구간길이가 같은 노출변수(exposure) 뿐만 아니라 기하구조 정보, 교통안전시설물 설치 정보, 도로 운영정보 등이 모형식의 독립변수로 사용된다. 통합 안전성능함수는 식(2)과 같이

표현될 수 있다(Korea Expressway Corporation, 2014).

$$N_{SPF} = e^{\alpha} \times ADT^{\beta} \times V_1^{\beta_1} \times V_2^{\beta_2} \times V_3^{\beta_3} \dots \times V_n^{\beta_n} \quad (2)$$

여기서,  $N_{SPF}$  = 예측사고건수 또는 예측사고율(건/km/년)  
 $ADT$  = 해당 구간의 일평균 교통량(대/일)  
 $\alpha, \beta_n$  = 회귀 계수  
 $V_1, V_n$  : 독립변수

### 2) 전통적인 회귀모형

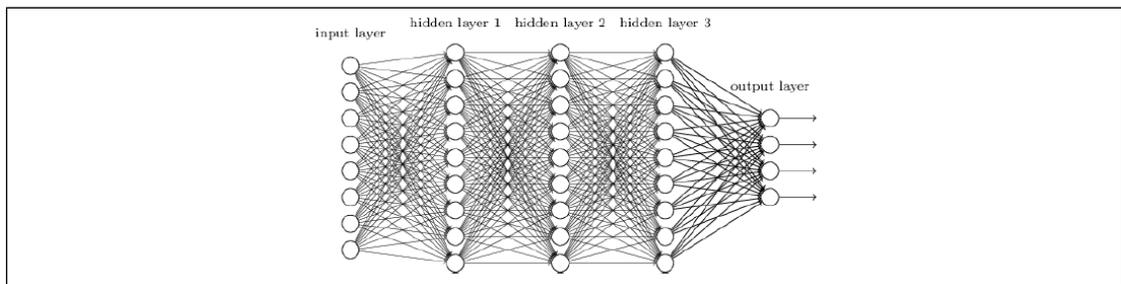
교통사고 건수와 같이 종속변수가 이산(discrete)적일 경우 포아송 회귀모형이 보편적으로 사용된다. 포아송 회귀모형을 교통사고 건수 예측에 사용하기 위해서는 종속변수인 교통사고 건수의 평균과 분산이 같다는 포아송 분포의 가정이 필요하다. 하지만 실제 교통사고 건수의 경우 이러한 가정을 만족하지 못해 과소산포(under-dispersion) 혹은 과대산포(over-dispersion)의 문제가 발생하는 경우가 있다. 음이항 회귀모형(negative Binomial regression Model)은 이러한 포아송 분포의 한계를 오차항( $\epsilon_k$ ) 추가를 통해 해결할 수 있다. 이때 기댓값은 아래 식(3)와 같이 표현할 수 있다(Simon et al., 2010; Jeong, 2014).

$$\mu_i = \exp(\beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \epsilon_k) = \exp\left(\sum_0^k \beta_k X_k + \epsilon_k\right) \quad (3)$$

여기서,  $\beta_k$  = 회귀식의 추정계수  
 $X_k$  = 회귀식의 독립변수  
 $\epsilon_k$  = 오차항

### 3) 딥 러닝

딥 러닝은 인공신경망(artificial neural network, ANN)이 가진 한계를 극복하기 위해 제안한 모형으로 심층 신경망, 합성곱 신경망, 제한 볼츠만머신, 심층 신뢰 신경망 등이 있다. 심층 신경망(deep neural network, DNN)은 기존 인공신경망과 구조가 같으며, 은닉층(hidden layer)의 개수가 2개 이상인 복잡한 모형이다. DNN 모형은 입력층(input layer), 은닉층, 출력층(output layer)으로 구성되어 있으며, 각 층은 노드로 구성되어 있다. 특히 노드의 개수가 많을수록 wide 모형이라고 칭하며, 은닉층이 많으면 deep 모형이라 부른다. 입력층의 노드 수는 입력 변수의 수와 같으며, 출력층의 노드 수는 회귀모형(regression)인 경우 1개로 구성되고, 분류 모형(classification)의 경우 분류할 개수가 된다. 은닉층의 노드 수는 사용자의 설계에 따라 달라진다. 다음 그림은 DNN의 구조를 나타낸 그림이다(Ryu, 2018).



〈Fig. 2〉 Deep learning network structure

## 2. 관련 연구 고찰

Seo et al.(2015)는 경부고속도로, 호남고속도로, 영동고속도로, 서해안고속도로, 중부내륙고속도로, 중앙고속도로를 대상으로 교통사고 특성을 분석하였으며, 고속도로 교통사고 건수 예측을 위한 안전성능함수를 개발하였다. 안전성능함수 개발에 사용된 변수는 교통사고 건수, 연평균일교통량(AADT), 구간길이이다. 안전성능함수 개발에 음이항 회귀모형을 A, B, C로 구분하여 함께 개발하였으며, 보정계수를 노선별로 개발하였다.

Jinyan et al.(2013)은 예측된 교통사고와 실제 교통사고 위험이 높은 위치를 검증하기 위해 단순 안전성능함수와 통합 안전성능함수를 개발하여 비교하였다. 성과지표(measures of effectiveness, MOE)로 중앙값 절대편차(median absolute deviation, MAD)와 평균제곱예측오차(Mean Squared Prediction Error, MSPE)를 사용하여 비교하였다. 비교 결과, 두 가지 안전성능함수 모두 유사한 성능을 보이는 것으로 확인되었다.

Guangyuan et al.(2017)은 다양한 지역의 다른 고속도로의 교통사고 빈도를 예측하는 데 사용할 수 있는 글로벌 안전성능함수(SPF)를 개발하기 위해 딥 러닝 기법을 적용하여 연구를 수행하였다. 딥 러닝 기법으로는 deep brief network(DBN)을 이용하여 기존 전통적인 통계모형인 음이항 회귀모형의 대안으로 제시하였다. 분석 결과 DBN 모형이 음이항 모형의 성능과 비교될 수 있음을 보여주었다.

Helai et al.(2016)는 고속도로 교통사고 예측모형을 radial basis function neural network(RBFNN) 모형을 이용하여 개발하였다. 또한 개발된 모형의 성능을 음이항 회귀모형, back-propagation neural network(BPNN)과 비교 분석하였다. 분석결과 RBFNN 모형이 음이항 회귀모형 및 BPNN 모델보다 높은 성능을 보이는 것으로 나타났다.

## Ⅲ. 자료 수집 및 구축

### 1. 변수 선정

교통사고는 인적, 물적, 환경적 요인이 복합적으로 작용하여 발생하는 것으로 관련 연구 고찰을 통해 고속도로 교통사고 건수 예측모형 구축을 위한 변수를 조사하였다. 대부분의 선행 연구에서는 고속도로 교통사고 건수 예측모형 개발을 위한 종속변수로 교통사고 건수를 사용하였다. 독립변수로는 교통량, 구간길이, 평면선형, 종단선형, 차로수, 제한속도 등 다양한 변수들이 사용되고 있었으며, 이들 중 공통적으로 사용되는 변수를 본 연구의 독립변수로 선정하였다.

### 2. 자료 수집 및 가공

선정된 변수를 바탕으로 고속도로 자료 수집을 위해 고속도로 구간을 분리한 콘존 단위로 자료를 수집하였다. 우선 종속변수인 교통사고 건수는 고속도로 교통사고 속도 자료를 2013~2015년까지 3년간 C급 이상 사고에 대해 수집하였다. 본 연구의 시간적 범위인 3개년(2013~2015년) 동안의 고속도로 전체 교통사고는 2013년 2,496건, 2014년 2,395건, 2015년 2,251건이 발생하였다. 교통사고로 인한 부상자는 2013년 1,253명, 2014년 1,148명, 2015년 1,054명이고, 사망자는 2013년 264명, 2014년 253명, 2015년 223명을 기록하였다(Ryu, 2018).

독립변수에 사용될 자료를 수집하기 위해서 2013~2015년 고속도로 교통량 자료 및 한국도로공사 ICT 센터가 보유한 교통통합 DB의 줄음 쉼터 위치, 휴게소 위치, 교량 위치, 기하구조 자료 등을 수집하였다. 수집된 자료들을 콘존 단위로 가공하기 위해 콘존 마스터 테이블의 콘존 ID를 기준으로 콘존 마스터와 매칭을

하였다. 하지만 교통사고 속보 자료, 졸음쉼터 자료, 휴게소 자료와 같이 콘존 ID가 없는 경우 노선명, 기종점방향, 이정을 이용하여 매칭하였다. 또한, 교량 위치, 시설물 위치, 터널 위치 자료의 경우 GPS 좌표를 이용하여 콘존 SHP 파일과 공간연산을 수행하였다. 공간연산은 대표적인 GIS 프로그램인 ArcGIS ver 10.3을 이용하였다. 분석 테이블을 생성하기 위해 콘존 ID가 매칭된 자료들을 콘존 마스터 테이블을 기준으로 분석 테이블을 구축하였다. 분석 테이블을 구축한 결과 종속변수와 독립변수를 포함하여 총 16개의 변수로 구성되었으며, 총 978개 데이터로 구축되었다. 분석테이블은 콘존 길이, AADT, 졸음쉼터개수 등 독립변수를 먼저 작성하고 종속변수인 사고건수를 마지막에 작성하였다. <Table 1>에서 보인 바와 같이 구축된 분석테이블의 16개 변수의 특징을 파악하기 위해 기초 통계분석을 실시하였다(Ryu, 2018).

<Table 1> Result of basic statistics for all variables

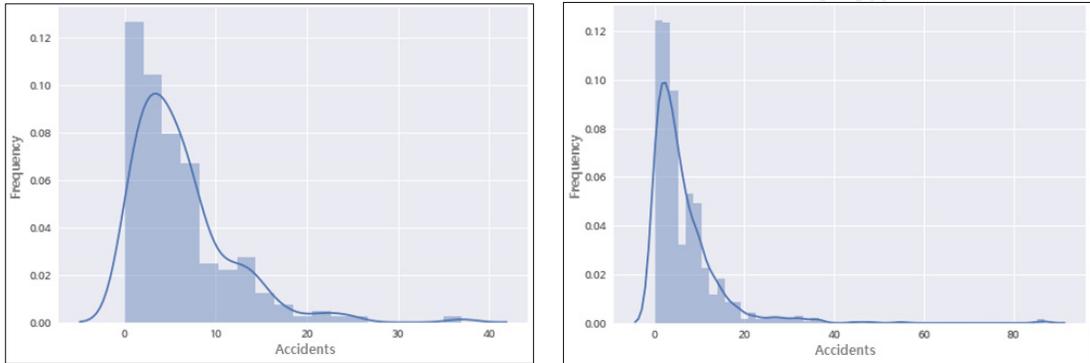
Variable	Mean	Standard deviation	Min	Max
Accident	6.39	7.49	0.00	87.00
Conzone length(m)	7,379.37	5,408.78	110.00	30,790.00
AADT	31,948.58	26,720.75	542.00	118,601.00
The number of drowsy rest areas	0.11	0.31	0.00	1.00
The number of rest areas	0.18	0.39	0.00	2.00
Design speed	103.27	8.81	80.00	120.00
Limited speed	101.02	6.56	70.00	110.00
The number of lanes	2.56	0.83	2.00	6.00
Side lanes	0.03	0.18	0.00	1.00
Entrance control segment	0.00	0.06	0.00	1.00
Bus-only lane	0.04	0.20	0.00	1.00
The number of tunnels	0.15	0.95	0.00	15.00
The number of bridges	4.68	4.21	0.00	23.00
The number of facilities	5.47	6.85	0.00	41.00
Curve radius	851.50	1,264.14	0.00	27,547.27
Vertical grade	0.01	0.54	-2.31	3.01

## IV. 모형 검증 방법

### 1. 학습 및 검증 데이터 분리

전통적인 통계 방법의 고속도로 교통사고 예측모형 및 딥 러닝을 이용한 고속도로 교통사고 예측모형 구축에 앞서 모형을 구축하는 학습데이터와 예측력을 검증하는 테스트 데이터로 구분하였다. 학습 데이터는 전체 데이터의 80%, 테스트 데이터는 20%로 랜덤하게 선정하였다. 학습 데이터와 테스트 데이터를 구분하기 위해 데이터 분석에 사용되는 프로그래밍 언어인 파이썬의 Scikit-learn 라이브러리를 이용하여 랜덤 샘플링 하였다. 랜덤 샘플링 결과 학습 데이터는 781건이고, 테스트 데이터는 196건으로 구분되었다. 또한, <Fig. 3>에서 보인 바와 같이 학습 데이터와 테스트 데이터의 분포는 동일한 것으로 파악되었다. 이렇게 학습 데이터와 테스트 데이터로 구분한 이유는 기존에 교통사고 건수 예측 등에 많이 활용되어온 인공신경망과 같은 기법들이 과대 추정(over-fitting) 문제로 인하여 학습 데이터에 대한 추정력은 매우 강하나 다른 지점 또는

다른 시간대와 같이 학습에 사용되지 않은 자료들에 대한 추정력에 문제를 보인 경험을 기반으로 교통사고 건수 예측모형에 새롭게 사용되는 딥 러닝도 이러한 문제를 보이는 지 확인하기 위해서이다.



〈Fig. 3〉 Accident distribution for learning data set 〈Fig. 4〉 Accident distribution for test data set

## 2. 모형 검증 방법

학습 데이터를 통해 구축된 모형을 검증하기 위해서는 테스트 데이터를 이용하여 모형의 예측력을 검증하는 것이 필수적이다. 모형의 예측력을 검증하는 방법으로는 MAD와 평균 제곱근 편차(Root Mean Square Error, RMSE), SMAPE를 사용하였다. MAD는 실제 사고건수와 모형을 통해 예측된 사고건수를 비교하기 위한 방법으로 실제 사고건수와 예측된 사고건수의 차의 절댓값을 구하여 산술평균한 것을 의미한다. MAD 값이 작을수록 모형의 설명력이 높다고 판단할 수 있다. RMSE는 고속도로 교통사고 예측모형을 통해 예측된 값과 실제 사고 건수의 차이를 다루는 측도로서 정밀도를 표현하는데 적합하다. 각각의 차이는 잔차라고 하며, 평균제곱근 잔차들을 하나의 측도로 종합할 때 사용된다. SMAPE는 통계적 기법으로 예측한 값에 대하여 정확도를 측정하는 방법인 기존 MAPE의 단점을 보완한 방법으로 백분율 또는 상대 오류를 기반으로 하는 정확도 측정 방법이다(Ryu, 2018).

## V. 교통사고 예측모형 개발 및 비교

### 1. 음이항 회귀모형을 이용한 고속도로 교통사고 예측모형 개발

본 연구에서는 고속도로 교통사고 예측모형 선정을 위해 수집된 자료를 이용하여 포아송 회귀모형과 음이항 회귀모형을 구축하여 우도비 검정통계량(Likelihood Ratio, LR)을 바탕으로 과분산을 검정하여야 한다. 검정결과 과분산이 있는 것으로 판정하여 최종적으로 음이항 회귀모형을 선정하였다(Ryu, 2018).

음이항 회귀모형을 이용한 고속도로 교통사고 예측모형 구축을 위해 데이터분석, 통계분석, 머신 러닝, 딥 러닝 구현이 우수한 프로그래밍 언어인 Python 3.6버전과 일반선형화 모형 라이브러리인 Statsmodel을 이용하여 음이항 회귀모형을 구축하였다. 모형은 AADT 및 구간길이를 노출계수로 하는 모형을 선정하여 사용하였으며, 모형식은 다음과 같다.

$$Y_B = AADT^{\beta_1} \times l^{\beta_2} \times e^{(\alpha + \beta_3 \dots)} \tag{4}$$

여기서,  $Y_B$  = 교통사고예측건수(건)  
 $AADT$  = 해당 구간의 연평균일교통량(대/일)  
 $l$  = 콘존 길이  
 $\alpha, \beta$  = 회귀 계수

<Table 2> Estimation result of negative binomial regression model

Variables	Coefficient	Standard Error	p-value
Constant	-13.704***	0.745	0.000
Conzone length	0.894***	0.048	0.000
AADT	0.812***	0.055	0.000
The number of rest areas	0.144*	0.080	0.070
The number of lanes	-0.269***	0.053	0.000
Bus-only lane	0.460***	0.156	0.003
The number of bridges	0.017**	0.008	0.034
Summary Statistics			
Number of observations : 781			
Number of variables : 6			
Log likelihood : -2077.2			
AIC : 4168.4			
BIC : -4751.8			

Note: \*\*\* significance level 0.01, \*\* significance level 0.05, \* significance level 0.1

모형 구축 결과, 유의수준 0.1에서 휴게소 수가 유의한 것으로 나타났으며, 유의수준 0.05에서 교량 수가 유의한 것으로 나타났다. 또한, 유의수준 0.01에서 콘존 길이, AADT, 차로수, 버스전용차로 여부가 유의한 것으로 나타났다. 특히, 모형 부호 검증 결과 차로 수가 적을수록 교통사고가 증가하는 것으로 나타났으며, 데이터 분석 결과, 적은 차로 수에서 높은 교통사고율을 보이는 것으로 나타났다. 최종적으로는 총 여섯 개의 변수들이 음이항 회귀모형에 포함되었다. 그 중에서 AADT와 콘존 길이는 노출계수로 사용되었다.

## 2. 딥 러닝을 이용한 고속도로 교통사고 건수 예측모형 개발

딥 러닝을 이용한 고속도로 교통사고 건수 예측모형을 개발하기 위해 딥 러닝 기법 중 다양하게 사용되고 있는 DNN을 이용하였다. 모형 구축을 위해 프로그래밍 언어인 Python ver 3.6과 딥 러닝 구현 라이브러리인 Tensorflow ver 1.4를 이용하여 모형을 구축하였다. DNN 모형 구축을 위한 변수는 앞서 음이항 회귀모형에서 사용된 변수인 AADT, 콘존길이, 휴게소 개수, 차로수, 버스 전용차로 여부, 교량 개수를 사용하였다. DNN 모형 구축을 위해서는 네트워크 구조 시나리오를 설계하는 것이 필요하며, 은닉층 수와 노드(node) 수를 결정하기 위해 민감도 분석을 수행하였다. 은닉층의 경우 처리 속도를 고려하여 3개와 5개를 고려하였고, 노드 수는 15, 25, 50, 75, 100을 이용하였다. 학습과 테스트를 5번 반복 수행하여 학습 데이터와 테스트 데이터의 cost가 적은 모형인 은닉층 3개와 노드 25개를 갖는 모형이 선정되었다.

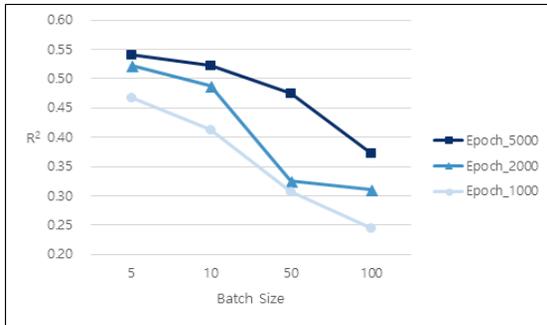
〈Table 3〉 Learning and test cost for hidden layer and node scenarios

Hidden Layer	3					5				
	15	25	50	75	100	15	25	50	75	100
학습	3.35	2.87	2.65	2.46	2.48	2.85	2.66	2.37	2.17	1.88
테스트	2.77	2.86	2.88	2.83	2.85	2.83	2.87	2.92	2.97	3.05
차이	-0.58	-0.01	0.23	0.37	0.37	-0.02	0.20	0.55	0.79	1.17

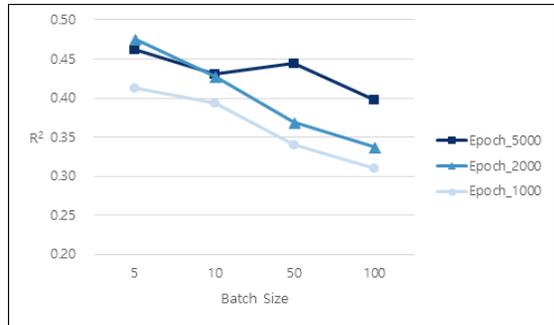
또한, 구축된 모형의 성능을 높이기 위해 비용 함수(cost function), optimizer, 배치 사이즈, epoch 등의 파라미터를 튜닝 하였다. 비용 함수로는 MAD를 선정하였으며, 학습 데이터와 테스트 데이터를 평가하는데 사용되었다. Optimizer로는 Adagrad를 선정하였다. 학습률(learning rate)을 조절해가면서 학습을 진행하는 최적화 기법으로 학습률을 초기에는 크게 했다가 점차 줄여가는 방법을 적용하였으며, 최적의 값을 빠르게 찾을 수 있다는 장점이 있다. Batch size 및 epoch는 민감도 분석을 통해 모형이 과적합되지 않도록 선정하였다.

〈Table 4〉 Parameter sets for deep learning model

Parameters	Values	note
Cost function	Mean absolute deviation	
Optimizer	Adagrad	
Batch size	5	
Epoch	5,000	



〈Fig. 5〉 R<sup>2</sup> by batch size in learning data

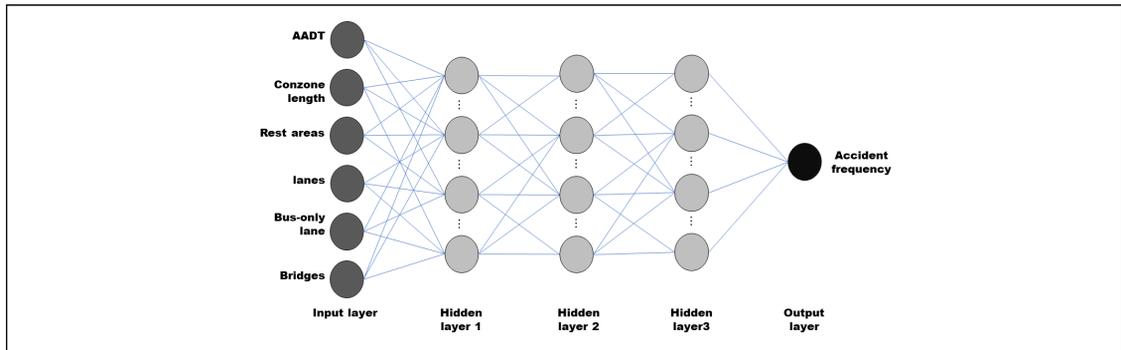


〈Fig. 6〉 R<sup>2</sup> by batch size in test data

또한 최종 모형을 선정하기 위해 앞서 선정된 은닉층 3개와 노드 25개를 갖는 함수를 기준으로 노드 구조 변경을 통해 최종 모형을 선정하게 되었으며, 학습 cost와 테스트 cost의 차이가 적은 것을 선정하였다. 선정 결과, 25-15-6 모형의 학습 cost와 테스트 cost의 차이가 0.07로 나타나 최종 모형으로 선정하였다.

<Table 5> Node structure selection result

Node Scenario			Learning cost	Test cost	Difference	Selection
Hidden layer1	Hidden layer2	Hidden layer3				
25	25	25	3.07	2.78	0.29	
25	50	25	3.00	2.85	0.15	
25	12	25	3.15	2.82	0.33	
50	25	12	3.00	2.84	0.16	
12	25	50	3.27	2.86	0.41	
25	15	6	2.59	2.52	0.07	Selected



<Fig. 7> Deep learning network structure

### 3. 모형 성능 비교

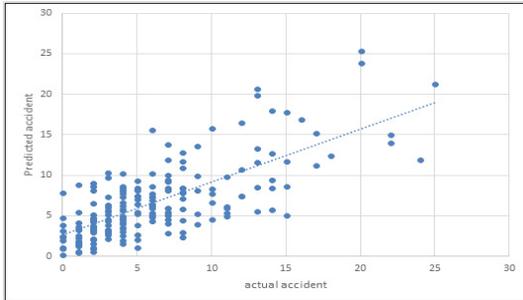
음이항 회귀모형과 딥 러닝으로 구축된 두 모형을 테스트 데이터를 이용하여 성능을 비교하였다. <Table 6>에서 보인 바와 같이, 딥 러닝을 이용한 교통사고 건수 예측모형이 MAD, RMSE, SMAPE 모두에서 우수하게 나타났다. 참고로 MAD, RMSE, SMAPE 값이 낮은 모형이 우수하다고 할 수 있다. MAD의 경우 2.52로 음이항 회귀모형에 비해 딥 러닝 모형이 0.27 낮게 나타났으며, RMSE의 경우 3.43으로 음이항 회귀모형에 비해 딥 러닝 모형이 0.24 낮게 나타났다. SMAPE의 경우도 딥 러닝 모형이 0.01 더 낮게 나타났다. 전반적으로 딥 러닝 모형의 성능이 음이항 회귀모형에 비해 우수하나 차이는 미미한 것으로 나타났으며, 이는 데이터 수의 한계로 판단된다.

<Table 6> Comparison result between negative binomial and deep learning models

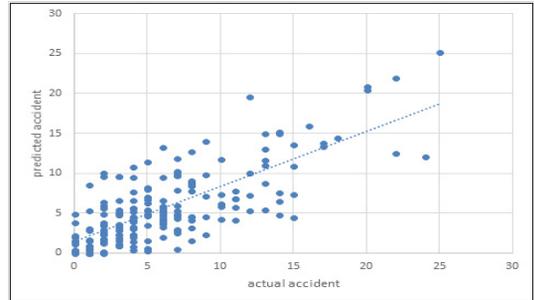
Model	MAD	RMSE	SMAPE
Negative binomial model	2.79	3.67	0.22
Deep learning model	2.52	3.43	0.21
Comparison	0.27	0.24	0.01

또한, 음이항 회귀모형과 딥 러닝 모형의 실제 사고건수와 예측 사고건수를 비교하기 위해 테스트 데이터의 실제 사고건수와 예측 사고건수를 시각화 하여 비교하였다. 비교 결과, 딥 러닝 모형의 경우 음이항 회귀

모형에 비해 실제 사고건수가 적은 구간에서 예측을 잘하는 것으로 나타났다. 하지만, 두 모형의 차이는 미미한 것으로 나타났다.



〈Fig. 8〉 prediction result by negative binomial



〈Fig. 9〉 prediction result by deep learning

## VI. 결론 및 향후 연구과제

### 1. 결 론

본 연구는 교통사고 건수 예측 분야에 전통적인 통계모형 대신에 최근 다양하게 활용되고 있는 딥러닝을 접목시켜서 그 적용성을 확인하기 위해 시작되었다.

기존에는 대부분의 교통사고 자료 분석과 교통사고 건수 예측 등이 전통적인 통계적 방법인 음이항 회귀모형을 이용하여 수행돼 왔다. 이러한 통계적 방법은 교통사고와 관련된 도로 및 환경 요소들과 교통사고 간의 인과관계를 찾고, 교통사고 빈도를 예측하는 방식으로 사용되었다. 하지만, 최근 머신 러닝 및 딥 러닝과 같은 분석 기법을 활용한 접근 방법들이 교통 및 다양한 분야에 적용되기 시작하였다. 이러한 머신 러닝 및 딥 러닝 기법은 다양하며 대량의 자료를 활용할 수 있는 장점이 있어 교통 및 다른 분야에서 활발하게 적용되어 우리들의 일상을 변화시키고 있다. 이에 본 연구에서는 고속도로 교통사고 자료를 이용하여 고속도로 콘존의 교통사고 건수를 예측하기 위해 음이항 회귀모형과 딥 러닝 모형을 이용하여 모형의 성능을 비교하였다. 예측 성능 비교 결과, 딥 러닝 모형의 MOE들이 음이항 회귀모형에 비해 다소 우수한 것으로 나타났다. 하지만 MAD 기준으로 차이가 0.27로 차이는 미미한 것으로 나타났다.

### 2. 향후 연구과제

본 연구는 딥 러닝을 이용하여 고속도로 교통사고를 예측하는 모형을 개발하였으나 몇몇 한계가 존재한다. 우선 자료의 구축 단위인 고속도로 콘존은 지방부의 경우 구간길이가 길고 다양한 구간길이를 포함하고 있어 딥 러닝을 수행하는데 충분한 데이터 수를 확보하지 못하는 한계가 존재한다.

또한 본 연구는 다른 도로 유형보다 상대적으로 교통사고 건수 예측에 필요한 기본 자료가 정확하고 많은 고속도로를 공간적 배경으로 활용하였지만, 교통사고 예측에 활용될 수 있는 상세한 자료가 부족하여 모형 개발에 다양한 변수를 고려하지 못한 한계가 존재한다.

본 연구의 한계점을 해결하기 위해서는 첫째, 교통사고 건수 예측에 딥 러닝을 이용하기 위해 입력변수의 다양화 통해 모형을 고도화하는 것이 필요하다. 현재는 공간적 범위로서 고속도로 콘존을 대상으로 모형을

구축하였으나, 자세한 자료가 많고 구간길이가 일정한 VDS 존으로 확대하여 분석하는 것이 필요하다. 둘째, 입력 자료를 확대하는 것이다. 입력변수도 세부 기하구조, 교통량 및 속도 시계열 자료 등으로 확대할 수 있다. 현재는 단면적인 거시적 자료를 이용하여 모형을 구축하였으나, 교통량 또는 속도 시계열 자료들을 활용한다면 해당 도로의 교통 패턴에 대해 적극적인 고려가 가능할 것으로 판단된다. 또한 향후 딥 러닝 활용 등을 염두에 두고 교통사고 자료 및 기하구조 자료 등을 DB로 구축하여 관리하는 것이 필요할 것으로 판단된다. 마지막으로 전통적인 통계적 모형의 경우 독립변수를 선정하는 일반적인 절차가 구축되어 있다. 딥 러닝과 같은 모형의 경우에도 그러한 절차가 있으나 본 연구에서는 그 절차를 고려하지 않았다. 예를 들어, 딥러닝을 적용할 경우 특성 추출(feature extraction) 또는 특성 선택(feature selection)과 같은 방법이 존재하므로 이러한 방법론에 대한 적용성 검토가 향후 필요하며 또한 교통사고 건수 예측과 같은 응용 사례에 적합한 입력변수(즉, 독립변수) 선정 방법을 개발하는 것도 장기적으로 필요하다.

## ACKNOWLEDGEMENTS

이 논문은 2015년도 정부(교육과학기술부)의 재원으로 한국연구재단(NRF)의 기초연구사업 지원을 받아 수행된 것임(2015R1A1A1A05028008).

## REFERENCES

- Huang H., Zeng Q., Pei X., Wong S. C. and Xu P.(2016), “Predicting crash frequency using an optimised radial basis function neural network model,” *Transportmetrica A: Transport Science*, vol. 12, no 4, pp.330-345.
- Jung J.(2014), *Poisson regression and negative binomial regression fit for traffic accidents*, Master thesis, Korea University.
- Korea Expressway Corporation(2014), *Highway Safety Performance Accident Correction Factor Development Casebook*.
- Lu J., Haleem K., Alluri P. and Gan A.(2013), “Full versus simple safety performance functions: a comparison based on urban four-lane freeway interchange influence areas in Florida,” *Transportation Research Board Annual Meeting*.
- National Police Agency(2017), *Traffic Accident Statistics in 2016*.
- Oh Y. T. and Kang D. S.(2017), *Advanced Traffic Safety Engineering*.
- Pan G., Fu L. and Thakali L.(2017), “Development of a global road safety performance function using deep neural networks,” *International Journal of Transportation Science and Technology*, vol. 6, no. 3, pp.159-173.
- Ryu J.(2018), *Development of expressway traffic accident prediction model using deep learning*, Doctoral thesis, Ajou University.
- Seo I. K., Kang D. Y., Park J. J. and Park S. H.(2015), “Development of safety performance function based on expressway alignment homogeneous section,” *Journal of the Korean Society of Civil Engineers*, vol. 35, no. 2, pp.397-405.