LSTM을 이용한 교통사고 발생 패턴 예측

Forecasting of Traffic Accident Occurrence Pattern Using LSTM

노 유 진* · 배 상 훈**

* 주저자 : 도로교통공단 운전면허본부 면허관리처장
** 교신저자 : 부경대학교 공간정보시스템공학과 교수

You Jin Roh* · Sang Hoon Bae**

* Koroad, Ph. D. Candidate, Pukyong National Univ.
** Professor, Pukyong National Univ.
† Corresponding author : Sang hoon Bae, sbae@pknu.ac.kr

Vol.20 No.3(2021)
June, 2021
pp.59~73

pISSN 1738-0774
eISSN 2384-1729
https://doi.org/10.12815/kits.2021.20.3.59

Received 13 May 2021
Revised 1 June 2021
Accepted 24 June 2021

요 약

교통사고로 인한 많은 인명피해가 발생하고 있으나, 첨단 기술의 발전에도 불구하고 교통사고 발생은 줄어들지 않고 있다. 교통사고를 사전에 예방하기 위해서는 향후 사고가 어떻게 변화하여 갈 것인지를 정확하게 예측할 필요가 있다. 지금까지 교통사고 발생 빈도 예측은 주요 연구 분야가 아니었으며 주로 과거 일정 기간의 통계를 기반으로 전통적인 방법으로 미시적으로 분석되어 왔다. 최근 AI 기술이 교통사고 분야에 도입 되었음에도 불구하고 주로 교통 흐름 예측에 초점을 맞추고 있어, 본 연구에서는 2014년부터 2019년까지 국내에서 발생한 1,339,587건의 교통사고 기록을 시계열 데이터로 변환하고 AI 알고리즘 LSTM을 이용하여 연령별, 시간별 교통사고 발생 빈도를 예측하였다. 또한 코로나-19로 인한 교통 환경의 변화에 맞추어 예측값과 실제값을 비교 검증하였다. 향후 이러한 연구결과가 교통사고 예방의 정책개선으로 이어지고 사고 예방에 활용 될 것으로 기대된다.

핵심어 : 교통사고, LSTM, 예측모델, 딥러닝, 시계열 분석, 코로나-19

ABSTRACT

There are many lives lost due traffic accidents, and which have not decreased despite advances in technology. In order to prevent traffic accidents, it is necessary to accurately forecast how they will change in the future. Until now, traffic accident-frequency forecasting has not been a major research field, but has been analyzed microscopically by traditional methods, mainly based on statistics over a previous period of time. Despite the recent introduction of AI to the traffic accident field, the focus is mainly on forecasting traffic flow. This study converts into time series data the records from 1,339,587 traffic accidents that occurred in Korea from 2014 to 2019, and uses the AI algorithm to forecast the frequency of traffic accidents based on driver's age and time of day. In addition, the forecast values and the actual values were compared and verified based on changes in the traffic environment due to COVID-19. In the future, these research results are expected to lead to improvements in policies that prevent traffic accidents.

Key words : LSTM, Road Traffic Accident, Forecast, Time series, COVID-19

Ⅰ. 서 론

1. 연구의 배경 및 목적

교통사고는 한해 수천 명의 소중한 생명을 앗아가고 사회적 비용이 연간 25조 9,000억원에 달하는 등 국가 경쟁력 약화 원인 중 하나이다. 교통사고로 인한 장애인수가 13,191명(2016)으로 교통사고 장애인의 69.9%가 실직을 하고 50대의 경우 78.6%가 근로능력을 상실하고 그들의 41.3%가 이혼 또는 별거로 가족해체의 아픔을 겪고 있다[1].

특히, 2020년 사망자가 출생자를 넘어서는 데드크로스가 처음으로 발생하였고, 인구 감소 32,700명 중 교통사고 사망자가 3,081명으로 9.4%를 차지하여 교통사고로 인한 피해가 국가적으로 큰 재앙이 되고 있다. 따라서 교통사고 발생 빈도수를 줄여나가야 진정한 사람 중심의 교통안전 국가가 실현될 것이다. 그럼에도 불구하고 2020년 교통사고 발생건수는 209,654건, 교통사고로 인한 사망자가 3,081명, 부상자가 306,194명으로 교통사고로 인한 고통을 전 국민이 겪고 있다고 할 수 있다.

교통사고를 줄이기 위하여 정부에서는 도심 내 차량속도를 감소시키는 안전속도 5030 정책[2], 고령자 교통사고 예방을 위하여 고령자 면허 적성검사 주기를 5년에서 3년으로 단축, 2019년 '윤창호법'시행으로 음주운전에 강력 대응, 어린이 보호를 위하여 2020년 '민식이법'을 제정하여 시행하고 있다. 하지만, 교통사고 발생 빈도수는 연간 20만건 이상으로 코로나-19 상황에서도 여전히 제자리걸음을 걷고 있다.

교통사고는 대부분 인적요인에 의해 발생하고 있지만, 기존의 연구는 도로환경적 요인, 차량적 요인을 포아송 모형 또는 음이항 회귀모형을 이용하여 분석하여 설명함으로서, 인적요인에 대한 연구가 상대적으로 부족하였다. 인적요인에 대한 연구는 도로교통공단의 TAAS 시스템에서 빈도분석과 교차분석으로 설명되어져, 시계열 패턴에 대한 연구가 미흡하였다.

최근에는 인공지능에 대한 이해도가 높고 많은 분야에서 딥러닝에 대한 연구가 진행되고 있다. 본 연구에서는 딥러닝 알고리즘 중 시계열 자료를 분석하는데 최적의 모형인 LSTM[3] 교통사고 예측 모델을 개발하였다. 교통사고 발생의 패턴을 찾아내기 위하여 도로교통공단 TAAS 시스템을 통해 필요한 자료를 추출하였다. 그리고 교통사고 자료의 전처리 과정을 거쳐 데이터 마이닝(Data Mining) 기법으로 분석하고 교통사고 빈도수 예측모형을 개발하여 인적요인에 대한 시계열 패턴을 찾아내어 교통사고 발생 꼭지점을 찾아 맞춤형 정책개발 선정에 활용될 수 있도록 하는데 그 목적이 있다.

2. 연구의 범위 및 방법

본 연구는 연간 약 20만 건이 발생하는 교통사고 발생 빈도수를 분석하기 위하여 경찰청에서 2014년부터 2020년까지 7년간 전국에서 접수, 처리되어 수집되는 교통사고 자료 1,549,151건을 도로교통공단 TAAS 시스템을 통해 필요한 자료를 추출하였다. 그런 후에 추출된 모든 자료를 입력 데이터 셋으로 이용될 수 있도록 전처리 과정을 다음과 같이 수행하였다. 첫째, 미 지정된 범주에 대하여는 정정, 삭제 또는 변경한다. 둘째, 문자형 범주는 통계 분석을 위해서 숫자로 변환된다. 셋째, 일단 모든 범주가 숫자로 변환되면 교통사고 자료를 시계열 형태로 빈도수를 합산하였다. 여기서, 시간과 같은 시퀀스 기준은 모든 범주에 대하여 이루어진다. 이러한 전처리는 데이터 전문 모듈인 PANDAS[4]의

1) KOTI(2018), "Results of Investigation on Victims of Traffic Accidents and Improvement Plans for Victim Support System," National Assembly Traffic Safety Forum
2) 제한속도를 도시부 50km/h, 주택가 등 이면도로 30km/h으로 하향
3) LSTM : Long Short-Term Memory

GROUP-BY 함수에 의하여 수행되었다. 전처리 과정을 거친 교통사고 자료는 데이터마이닝을 통하여 분석된 교통사고 자료 중에서 시계열 패턴을 보여주는 연령대별, 시간대별 자료를 LSTM 모형으로 교통사고 발생 빈도를 예측하였다. 기존의 연구에서 제시하지 못했던 교통사고 발생의 연령대별, 시간대별 추이를 파악하고, 특정 연령대와 특정 시간대에서 교통사고를 예방하기 위한 대안을 제안하고자 하였다.

본 연구 과정에서 코로나-19로 인한 전 세계 인구 절반 이상이 자가 격리에 들어갔으며 이동 제한 조치를 발령한 시기로서, 자가용을 이용한 이동도 줄어들면서 교통사고로 인한 사망자와 중상자가 급격하게 감소하였다[5]. 우리나라의 경우 2020년에는 교통사고 발생 빈도는 8.7%, 사망자 수는 8.0%, 중상자 수는 4.4% 감소하는 등 코로나-19로 인한 영향이 나타나고 있다. 이러한 사회 현상을 고려하여 정형 데이터 예측 모델을 설명하고자 한다.

3. 기존 연구

전통적인 통계기법을 이용하여 교통사고 빈도수 예측을 한 사례로는, Lee et al.(2003)는 도로 신설 및 개량 사업에 대한 타당성 조사 시 도로의 물리적 특성이 충분히 반영되지 못하는 점을 해결하기 위하여 교통 특성과 도로의 물리적 특성을 고려한 교통사고 예측모형을 개발하였다. Lee and Roh(2015)는 서울, 수도권, 부산의 4지 교차로를 대상으로 음이항 회귀모형을 이용하여 교통사고 예측모형을 구축하였다. 개발된 모형을 이용하여 교통사고 빈도 및 특성을 분석한 결과, 기존의 음이항 회귀모형 보다 확률적 음이항 회귀모형의 설명력을 높게 평가하였다. 많은 연구들이 교통사고 자료를 분석하여 전통적으로 사고에 영향을 미치는 여러 요인들 간의 상관관계 분석을 통해 계량적으로 평가하는 방법을 적용하고 있었다. 이는 도로의 다른 모든 조건이 동일한 상황에서 일부 특정 요인을 변화시킴으로서 이로 인해 개선되는 교통사고 발생 요인들 간의 효과를 안전성능함수(Safety Performance Function, SPF)[6]를 이용해 조건의 효율성을 평가하는 방법이다. 교통사고 관련 데이터 분석은 TAAS 자료를 활용하여, 매년 8월 중 전년도의 교통사고 통계분석 자료를 도로교통공단에서 발행하고 있으며, 회귀모형 기법으로 교통사고 간의 상관관계 분석을 통해 교통사고 건수를 예측하는 연구가 대부분이었다.

최근 인공지능 기술의 발달과 다양하고 대용량의 빅 데이터를 결합하여 교통사고 발생을 예방하기 위한 연구가 활발하게 이루어지고 있다. 한국에서는 Oh et al.(2014)는 교통사고 예측모형 구축에 주로 사용되는 회귀모형, 인공신경망, 구조방정식을 이용하여 교통사고 빈도수 예측모형을 각각 개발하였다. Ryu(2018)은 딥 러닝 모형 중 DNN을 이용하여 고속도로 교통사고 예측 모형을 구축하여 전통적인 음이항 회귀분석과 비교하여 딥 러닝이 교통사고 관련 연구에 유용하다는 것을 제안하였다. Han(2019)은 딥 러닝 모델인 DeepFM을 개발하여 교통사고 발생 방송제보, 기후, 교통소통현황 그리고 공공3.0정보와의 인과관계를 전국의 약 4,500개 사고다발지점을 대상으로 분석하였다. 중국에서는 Honglei Ren et al.(2017)이 딥 러닝 모형을 이용하여 교통사고 발생에 가장 중요한 요인이 교통흐름에 있다고 분석하였다. Zhihao et al.(2020)은 LSTM-GBRT 기반으로 교통사고 예측 모델을 수립하고 관련 데이터를 훈련시켜 교통사고 안전 수준 지표를 예측하였다.

캐나다 WATERLOO 대학의 Guangyuan et al.(2017)은 딥 러닝 기법 중 Deep Brief Network(DBN)을 이용하여 기존 회귀모델의 대안으로 crash modeling를 구축하였다. 실제 충돌 데이터 세트를 사용하여 다양한 지역의 고속도로의 교통사고 빈도를 예측하였다. 또한, 통계모형인 음이항 회귀모형보다 딥 러닝 기법이 교통사고 빈도수 예측

4) 파이썬 언어로 작성된 수치형 테이블과 시계열 데이터를 조작하고 운영하기 위한 데이터를 제공하는 소프트웨어 라이브러리이다. https://namu.wiki/w/pandas

5) COVID-19 Transport Brief : Re-spacing Our Cities for Resilience

6) Safety Performance Function, SPF : 교통사고 빈도수 또는 심각도 등의 종속변수와 교통운영체계, 도로관리체계, 그리고 차량관리 등 관련 독립변수들 간의 관계를 수학적으로 표현한 것.

하는 대안으로 우수함을 증명하였다.

유럽에서는 Benoit(2019)이 Facebook Prophet과 Keras/Tensorflow를 사용하는 LSTM 신경망을 개발하여 스위스의 교통사고를 예측하고 시각화하였다. 말레이시아의 Maher et al.(2017)은 6년 동안 말레이시아 남북 고속도로(NSE)에서 발생한 교통사고 기록을 이용하여 기존의 신경망 (NN)과 비교하여 RNN 방법이 교통사고 부상 심각도를 예측하는데 유리하다고 분석하였다.

많은 연구에서 딥 러닝을 이용하여 교통사고 예측을 하였으며, 다른 알고리즘 모델보다도 RNN 계열이 시계열 예측에 최적화되었음을 증명하였다. 하지만 대부분의 연구가 교통사고 발생의 예측력을 높이기 위한 모형의 비교 분석에 집중하였으며 교통사고 발생을 억제하기 위한 방안 마련에는 미흡한 것으로 나타났다.

Ⅱ. 교통사고 발생 예측모델 개발

1. TAAS 데이터 전처리

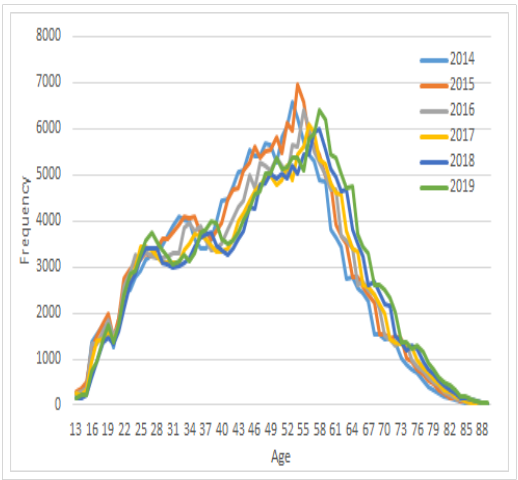
도로교통공단의 TAAS로부터 수집한 교통사고 데이터 1,549,151건을 전처리 과정을 거쳐 분석한 결과, 2014년~2016년 자료와 2017년~2020년 자료의 통계원표가 교통 환경의 변화를 반영하여 총 68개 항목에서 60개 항목으로 축소되었다. 본 연구에서는 통계원표 자료의 연속성 확보, 통계원표의 용어 통일 등으로 일부 자료를 활용하지 못하였다.

특히, 연령대 데이터에서는 12세 이하와 90세 이상의 데이터에서 오류 발생이 있어 교통사고 발생 가해자의 연령을 13세부터 89세까지 한정하여 처리하는 것이 정확한 분석을 위해 필요한 것으로 나타났다. 먼저 전처리 과정을 거치게 되면 다양한 교통사고 요인과 교통사고와의 관계 시각화가 가능하다. 시각화 과정을 거쳐 본 연구에서는 교통안전 정책과 수단에 유용한 근거가 될 수 있는 요인에 대한 교통사고 빈도수를 예측 대상으로 선정하였다.

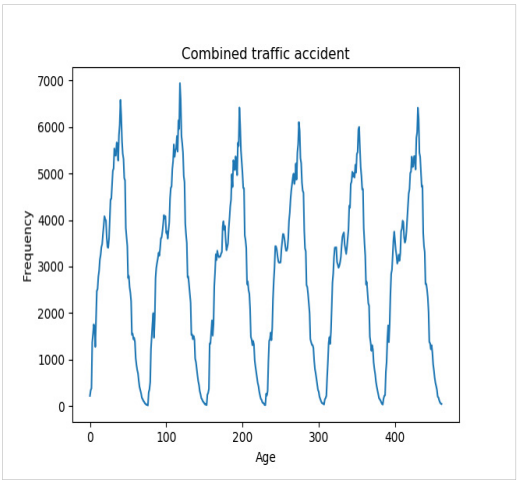
첫째는 연령대별 교통사고 빈도수 예측이다. 예를 들어, 가해자 13 ~ 89세 구간 나이 계열로 각 나이별 사고건수를 합산하여 할당하는 것이다. 그런 후에 증가하는 연도에 따라서 사고수를 연결하면 하나의 시계열 곡선 형태가 되는 것이다.

〈Fig. 1〉은 전처리된 연령별 교통사고 발생 빈도수를 보여주고 있다. 〈Fig. 1〉의 각 연도별 곡선을 〈Fig. 2〉와 같이 연결하면 시계열 추이가 되는 것이다. 각 연령대별 교통사고 유형이 다르고 교통사고를 가장 많이 유발하는 연령대를 분석하여 맞춤형 교통안전 정책 수립이 필요하며, 그리고 고령화 사회에 진입하는 한국의 경우, 고령자에 의한 교통사고 비중에 점차 증가하고 있어 그에 대한 이해와 대책이 중요하게 요구되기 때문이다. 교통사고를 가장 많이 유발하는 연령대와 연도별 추세를 파악하여 정책 수립에 활용될 수 있을 것이다.

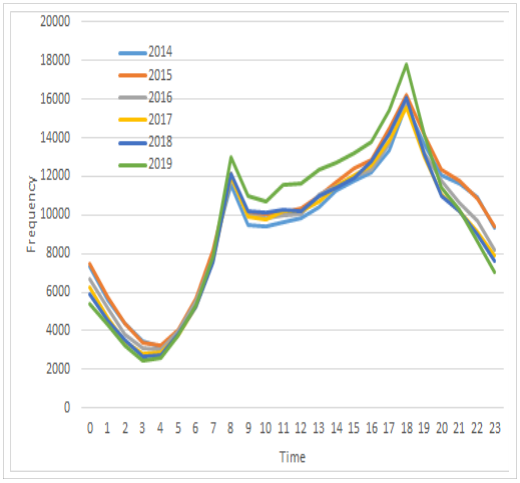
둘째는 시간대별 교통사고 빈도수를 선정하였다. 〈Fig. 3〉는 전처리된 시간대별 교통사고 발생 빈도수를 보여주고 있다. 〈Fig. 3〉의 각 연도별 곡선을 〈Fig. 4〉와 같이 연결하면 시계열 형태가 되는 것이다. 운전자들의 일상생활 패턴에 따라 교통사고 빈도수도 변화한다는 기본적인 전제에, 실제 교통사고 발생이 교통량이 급증하는 출퇴근시간대에 집중되는지를 파악하고 가장 많은 교통사고를 유발하는 시간대를 분석하여 시간대별 맞춤형 교통정책을 수립할 수 있다. 이러한 시간대에 따른 일상생활 패턴이 연도에 따라서 어떤 추세를 형성하는지도 교통사고를 줄이는 교통정책에 반영될 필요가 있다.



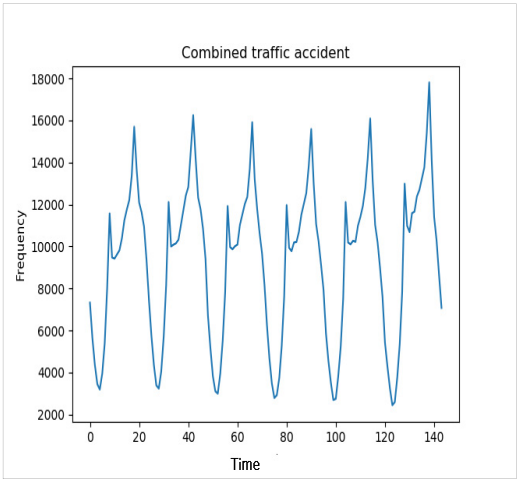
<Fig. 1> Frequency of traffic accidents by age



<Fig. 2> Time-series of traffic accidents by age



<Fig. 3> Frequency of traffic accidents by time



<Fig. 4> Time-series of traffic accidents by time

2. 데이터 마이닝 수행

본 연구에서 교통사고 데이터는 사고 당시 상황을 기록한 데이터로 운전자, 사고장소, 사고 종류, 차량 정보, 법규 위반 사항 등 다양한 속성으로 이루어져 있다. 데이터 마이닝은 이러한 데이터로부터 잘 알려져 있지 않은 유용한 패턴을 찾는 과정이다. 패턴은 규칙, 의사결정 트리, 특정 데이터의 집합과 같이 여러 가지 형태로 표현될 수 있으며 데이터의 특성이나 속성들 사이의 관계에 대한 정보를 제공한다. 어떤 패턴이 유용하다는 것은 패턴에 담긴 정보가 새롭고 미래에 유용하게 활용 가능하다는 것을 뜻한다.

경찰청의 교통사고 통계원표의 교통사고 개요 1~13개 항목, 당사자 14~30개 항목, 차량 31~37개 항목, 도로환경 38~51개 항목, 기타 피해자 52~60개 항목에 대하여 각 속성의 전처리와 더불어 데이터의 일부 필드에 대해 그룹

화나 이산화과정을 수행하였다. 통계원표에서 알 수 있듯이 교통사고 데이터 항목에 인적요인이 17개 항목, 차량요인이 7개 항목, 도로환경 요인이 14개 항목으로 구성되어 있다. 각 경찰서별로 교통사고 자료가 기록되는 과정에서 왜곡이 발생하여 전체 항목에서 데이터 내에서 어떻게 사용되는지 확인하는 과정이 필요하다. 아래 〈Table 1〉은 통계원표 자료를 TAAS 시스템에서 추출하여 정리한 변수들의 집합이다.

<Table 1> Data field of Traffic Accident (2019)

1																	
2	2019																
3																	
4	사고일시	사고시간	요일	시도	시군구	사고내용	사망자수	중상자수	경상자수	부상신고자수	사고유형		법규위반1당	노면상태	기상상태	도로	
5	2019010100	00시	화	강원	원주시	경상	0	0	1	0	차대차	추돌	안전운전 의	포장	건조	맑음	단일로 기
6	2019010100	00시	화	경기	남양주시	중상	0	2	0	0	차대사람	횡단중	보행자 보호	포장	건조	맑음	기타 기
7	2019010100	00시	화	경기	수원시	경상	0	0	1	0	차대차	측면충돌	안전운전 의	포장	건조	흐림	교차로 교
8	2019010100	00시	화	경기	양주시	중상	0	1	0	0	차대차	측면충돌	신호위반	포장	건조	맑음	교차로 교
9	2019010100	00시	화	경기	포천시	경상	0	0	1	0	차대차	기타	직진 및 우회	포장	건조	맑음	교차로 교
10	2019010100	00시	화	경기	화성시	중상	0	1	0	0	차대차	추돌	안전운전 의	포장	건조	맑음	단일로 기
11	2019010100	00시	화	광주	광산구	경상	0	0	1	0	차대차	기타	중앙선 침범	포장	건조	맑음	단일로 기
12	2019010100	00시	화	대구	달서구	경상	0	0	1	0	차대사람	보도통행중	기타	포장	건조	흐림	단일로 기
13	2019010100	00시	화	대구	중구	중상	0	1	0	0	차대사람	길가장자리구역	안전운전 의	포장	건조	맑음	단일로 기
14	2019010100	00시	화	부산	남구	중상	0	1	0	0	차대차	추돌	안전운전 의	포장	건조	맑음	교차로 교
15	2019010100	00시	화	부산	해운대구	사망	1	0	0	0	차대사람	기타	중앙선 침범	포장	건조	맑음	교차로 교
16	2019010100	00시	화	서울	강서구	경상	0	0	1	0	차대사람	횡단중	보행자 보호	포장	건조	맑음	교차로 교
17	2019010100	00시	화	서울	구로구	경상	0	0	1	0	차대차	추돌	안전운전 의	포장	건조	맑음	단일로 기
18	2019010100	00시	화	서울	서초구	경상	0	0	1	0	차대차	기타	안전운전 의	포장	건조	맑음	기타 기
19	2019010100	00시	화	인천	부평구	경상	0	0	1	0	차대사람	길가장자리구역	안전운전 의	포장	건조	맑음	교차로 교
20	2019010100	00시	화	인천	부평구	경상	0	0	2	1	차대차	측면충돌	신호위반	포장	건조	맑음	교차로 교

이러한 데이터마이닝을 통하여 분석된 교통사고 자료 중에서 시계열 패턴을 보여주는 연령대별, 시간대별 자료를 LSTM 모형으로 미래의 교통사고 발생 빈도를 예측하였다. 기존의 연구에서 제시하지 못했던 교통사고 발생의 연령대별, 시간대별 추이를 파악하고, 특정 연령대와 특정 시간대에서 교통사고를 예방하기 위한 대안을 제안하고자 하였다.

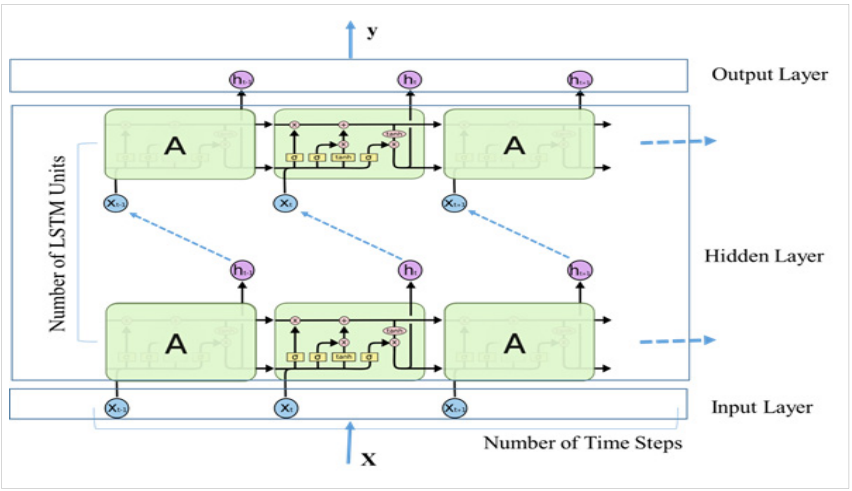
3. LSTM 관련 이론 고찰

교통사고 요인들에 따른 발생 빈도수를 예측하기 위해서는 사고 빈도수 패턴을 먼저 검토할 필요가 있다. 연령대별, 시간대별 교통사고 발생 빈도수는 〈Fig. 1〉 ~ 〈Fig. 4〉에서와 같이 시계열적이고 일정한 패턴을 가지면서 미래에는 과거의 추세를 따라간다고 판단할 수 있다. 즉, 교통사고 발생 빈도수는 주기(period), 추세(trend), 계절 효과(seasonality) 특성을 갖는 시계열 경향을 포함하는 것을 알 수 있다. 그렇기에 인공지능 알고리즘 중 시계열의 예측에 널리 이용되는 LSTM을 도입하게 되었다. 교통사고 발생 빈도수 중에서 연령대별, 시간대별 패턴은 주기성을 갖으면서 파형은 점차적으로 변하는 것을 볼 수 있다.

본 연구에서는 하나의 은닉층(hidden layer)과 출력을 예측하기 위한 하나의 출력층(output layer)을 갖는 LSTM 모델을 사용하였다. 즉, 〈Fig. 5〉에서 Hidden layer = 1, Output layer = 1인 LSTM이며, 이 모델을 사용한 이유는 본 논문에서 다루는 시계열 데이터가 비교적 단순하기 때문이다. 은닉층 내의 LSTM Unit 수와 입력 데이터의 Time Step 수는 최적화된다. 출력층은 Dense layer로 구성된다. LSTM Unit의 활성함수(activation)는 ReLU[7]가 이용되었다. LSTM 모델의 학습과 정확도에 영향을 미치는 입력변수들은 은닉 층의 레이어 수 및 각 은닉

7) ReLU : Rectified Linear Unit 입력값이 0보다 작으면 0이고, 0보다 크면 입력값 그대로를 내보낸다.

층에서의 노드 수, 활성화 함수, 최적화 기법, 한 번에 학습할 데이터의 규모와 학습을 얼마나 반복해야 할지 등이 있다. 본 연구에서의 진행 사항은 다음과 같다.



<Fig. 5> LSTM unit

- 은닉층의 수, 은닉층의 노드 수 : 은닉 층의 수는 1개이며, 은닉 층에서의 노드 수는 연령대별 빈도수 예측에서는 76개로, 시간대별 빈도수 예측에서는 24개로 시행착오법[8]을 통해 결정하였다.
- activation function : 활성화 함수는 모델의 출력을 결정하며, 가중치에 의한 연산 결과를 다음 층으로 전달하는 역할을 수행한다. 본 논문의 훈련데이터는 차원이 크지 않아 활성화 함수의 경우 시그모이드 함수의 변형이며, 일반적으로 널리 활용되는 쌍곡탄전트(hyperbolic tangent)를 활성화 함수로 설정하였다.
- loss function : Mean Squared error (MSE) 방법을 사용하였으며, 예측값에 대한 오류를 숫자로 나타낸 것으로 오류가 클수록 큰 값이 나오고 반대로 오류가 적을수록 작은 값이 나온다.
- optimization technique : 연산 메모리의 부하를 적게 하는 장점이 있는 Adam 방법을 사용하였으며 Adam 최적화기의 입력 매개변수(파라메터)는 Keras의 기본 값을 이용한다.
- batch size : 데이터의 크기에 따라 훈련데이터를 한 번에 학습하기 어려울 수 있으므로 훈련 데이터 셋을 나누어서 한다. 즉, 매 훈련 단계마다 학습할 데이터의 크기를 결정하는 항목이며 본 논문에서는 시행착오법을 통하여 각각 연령별 예측에서는 76개, 시간대별 예측에서는 24개로 결정하였다.
- epoch : 흔히 세대로 번역되는 epoch는 배치 사이즈로 분할된 훈련 데이터 전체를 1회 학습하는 횟수를 몇 번 반복할지를 의미한다.

LSTM 모형의 성능 최적화를 위해서는 알고리즘에 사용되는 최적의 하이퍼-파라메터 값을 찾는 것이 모형 설계에서 가장 중요하다. 하이퍼-파라메터 튜닝 과정은 정해진 방법이 없으며, 반복적인 실험과 시행착오를 거쳐 최적의 하이퍼-파라메터를 찾을 수 있다. 하이퍼-파라메터는 절대적으로 가장 좋은 값은 존재하지 않지만 사용하는 데이터와 모형에 따라 적합한 값을 찾을 수 있다.

8) 시행착오법 : 계속적으로 해답을 찾을 때까지 실제로 시험해 보면서 문제의 해답을 찾아가는 문제 해결 방식

Ⅲ. 예측 모델 개발

본 연구에서 연령대별, 시간대별 학습 데이터셋은 2014~2018년 교통사고 기록이고 2019년은 검증 데이터셋으로 이용된다. 2020년 교통사고 발생 빈도수를 예측하기 위하여 2개의 예측 대상 교통사고 빈도수 데이터에 대한 학습과 평가 결과를 〈Table 2〉에서 제시하고 있다. 학습과정에서 학습 데이터와 검증 데이터와 차이가 없는지 비교해 보니 비교적 큰 차이가 없는 모습을 보였고 검증 데이터에 과대적합 현상도 발생하지 않았다. 예측 모델의 정확도 평가는 개발자의 정성적 평가에 의한다. 즉, 연도별 교통사고 빈도수 분포의 주요 특징은 직관적으로 예상할 수 있다.

<Table 2> Optimal hyper-parameters for the LSTM model

Case	LSTM Unit	Time Steps	epoch	Loss function (MSE)	Metric (MAE)	Choice
Age	76	76	4,000	4.3031e-04	0.0154	
	76	152	4,000	2.5299e-04	0.0113	○
	152	76	4,000	4.2441e-04	0.0140	
	152	152	4,000	3.1742e-04	0.0139	
Time	24	24	4,000	5.0987e-04	0.0179	
	24	48	4,000	2.9960e-04	0.0134	○
	48	24	4,000	4.9751e-04	0.0163	
	48	48	4,000	2.8768e-04	0.0139	

본 연구에서는 LSTM 모델 개발을 위하여 어느 정도의 학습 능력을 가지고 있는지를 확인하기 위하여 손실함수(Loss function)를 사용하였으며, 예측한 값과 실제 사고 빈도수의 차이를 다루는 정밀도를 표현한 식 (1)의'평균제곱오차(Mean Square Error, MSE)'방법을 선정하였다. 평균제곱오차는 손실함수 중 가장 많이 쓰이며 예측값과 실제 값 사이의 평균을 제곱하여 평균을 낸 값이다.

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(y_i - t_i)^2 \quad \cdots\cdots (1)$$

그리고 LSTM의 정확도를 측정하기 위한 척도(Metric)로서 교통사고 발생 빈도수와 예측된 빈도수와의 차에 절댓값을 구하여 산술평균하는 식 (2)와 같이'절대평균오차(Mean Absolute Error, MAE)'방법을 사용하였다. 평균절대오차는 예측값과 실제 값의 차이에 절댓값을 평균한 값이다. 손실함수와 척도의 Error 값은 모두 그 값이 0에 가까울수록 높은 정확도를 나타낸다.

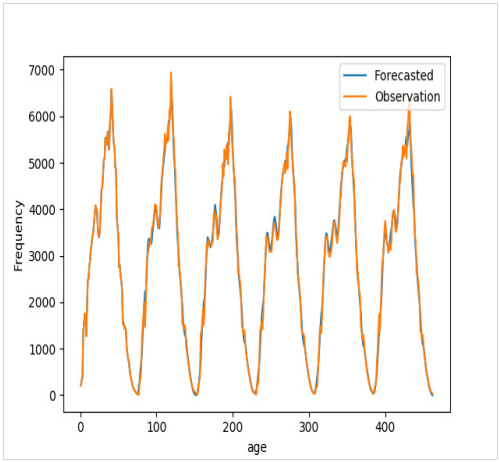
$$MAE = \frac{1}{N}\sum_{i=1}^{n}|y_i - t_i| \quad \cdots\cdots (2)$$

본 연구에서 최적화시켜야할 하이퍼-파라메터는 각 Hidden layer 내부의 Number of LSTM Unit, Number of Time Steps 그리고 Number of epoch 이다. 각각의 하이퍼-파라메터를 변경하면서 정확한 예측을 얻는다. epoch 수는 시행오차를 통하여 4000으로 선정하였다. 결과적으로 최적화시켜야할 매개변수는 2개만 남는다.

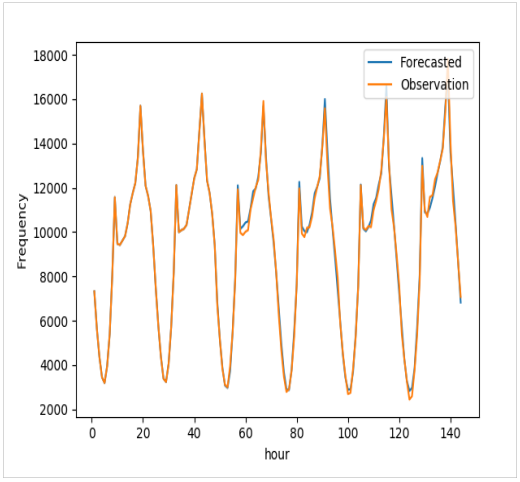
〈Fig. 6〉에서 연령대별 각 연도별 교통사고 발생 빈도 추세는 2014~2018년 학습 데이터 셋으로 2019년 검증한

결과값이다. 이 평가방법으로 하이퍼-파라메터를 최적화한 결과 연령대별 n_step=76, LSTM Unit=152 일 때, 직관적으로 수용할 수 있는 예측 결과를 얻었다.

〈Fig. 7〉에서 시간대별 각 연도별 교통사고 발생 빈도 추세는 2014~2018년 학습 데이터 셋으로 2019년 검증한 결과값이다. 이 평가방법으로 하이퍼-파라메터를 최적화한 결과 시간대별 n_step=24, LSTM Unit=48 일 때, 수용할 수 있는 예측 결과를 얻었다.



<Fig. 6> Training of traffic accidents by age



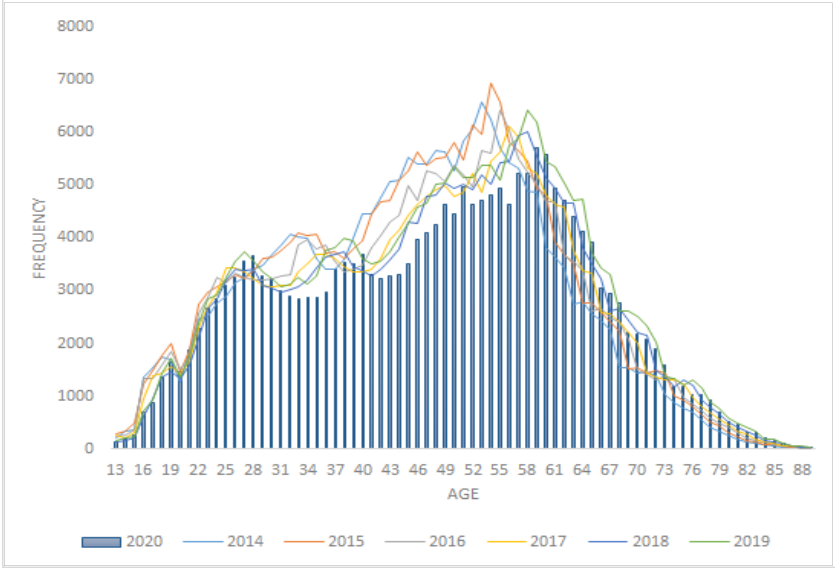
<Fig. 7> Training of traffic accidents by Time

Ⅳ. 예측 모델 분석 결과

1. 연령대별 예측모델

2020년 연령대별 교통사고 발생 빈도수를 예측하기 위하여 2014~2018년 학습 데이터 셋으로 2019년 검증한 결과값의 최적화한 하이퍼-파라메터를 이용하여 〈Fig. 8〉과 같이 직관적으로 수용할 수 있는 예측 결과를 얻었다. 〈Fig. 8〉에서 볼 수 있듯이 교통사고를 가장 많이 발생시키는 연령대는 50대 후반에서 60대 초반이다. 2014년에는 53세가 6,581건, 2015년에는 54세가 6,941건, 2016년에는 55세가 6,416건, 2017년에는 56세가 6,103건, 2018년에는 58세가 6,000건, 2019년에는 58세가 6,412건으로 매년 1년씩 연령이 증가하고 있으며, 태어난 연도로 환산하면 1961년 태어난 연령에서 가장 교통사고 가해자가 많다는 것을 알 수 있다. 2020년 예측에서는 59세에서 6,074건이 발생할 것으로 나타났다. 이러한 현상은 1960년~1963년 태어난 연령대가 가장 교통사고를 많이 발생하는 것으로 사회적 활동이 여전히 많다는 것을 보여주고 있다.

다음으로 교통사고가 꼭지점을 나타내는 연령대는 30대 초중반으로 2014년 32세가 4,078건으로 매년 1~2년씩 연령이 증가하면서 2019년에는 39세가 3,982건으로, 2020년 예측으로는 40세가 3,848건이다. 20대에서는 2014년에 22세가 2,473건으로 21세의 1,880건 보다 593건이 더 많이 발생하였으며, 매년 연령대가 높아지면서 2019년에는 27세에서 3,746건으로 증가하고 있다. 2020년에는 29세에 3,482건으로 예측되었다. 10대에서는 매년 19세가 가장 교통사고를 많이 발생시키고 있다. 운전면허를 처음 취득하는 연령이기도 하고, 초보운전자로서 2014년에는 1,690건에서 2019년 1,728건으로 증가하였고 2020년에는 1,865건으로 증가할 것으로 예측되었다.



<Fig. 8> forecasted of traffic accidents by Age

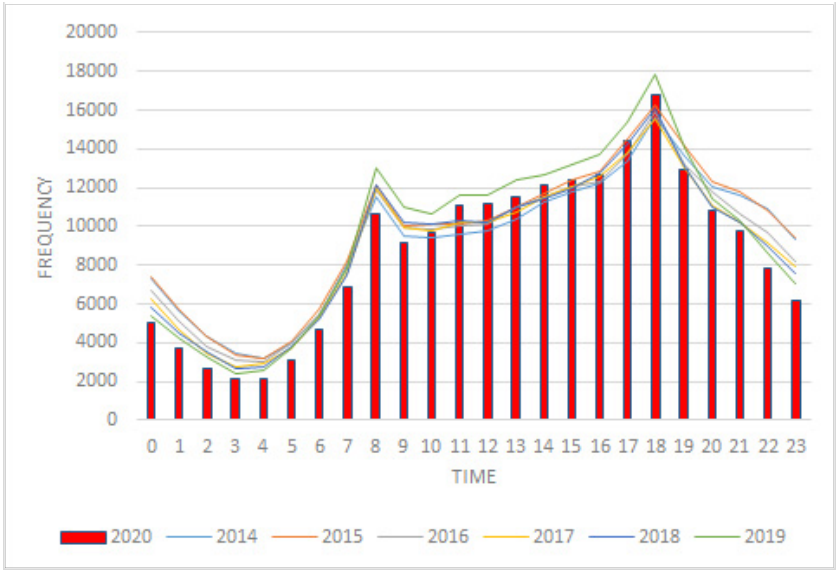
전체적으로 교통사고는 매년 연령대가 1~2년씩 높아지면서 특정 출생년도에서 교통사고를 많이 발생하는 것으로 분석된다. 이는 특정 출생년도가 경쟁이 심한 인구통계학적, 사회학적 요인이기도 하지만 여전히 사회적 활동을 하고 있다는 것으로 판단된다. 교통안전정책수립에서도 이러한 인구통계학적, 사회학적 관점에서도 관심을 가지고 연령대에 맞는 교통안전교육, 적성검사 기간 조정, 안전운전 재교육 등 다양한 프로그램이 개발되어야 할 것이다. 또한, 운전면허를 처음 취득하는 연령대에는 맞춤형 안전교육이 이루어져야 한다.

운전면허를 처음 취득하는 19세의 연령대는 매년 교통사고가 꼭지점을 찍는 연령대이다. 19세를 정점으로 교통사고 발생 빈도가 감소하고 있다. 특히 19세부터 20세 까지는 이륜차로 인한 교통사고가 가장 많이 발생하고 있다. 70~80대는 다른 연령층에 비해 교통사고 비율이 높지 않은 상황이지만, 본격적인 고령화 사회로 진입하면서 운전면허 적성검사 강화와 같은 대책이 필요하다. 대도시 고령 운전자에게는 운전면허반납제도의 활성화와 더불어 농어촌에 거주하는 고령운전자에게는 농기계에 대한 교통사고 예방활동도 병행하여야 한다. 지속적으로 사회활동의 연령대가 높아질수록 매년 교통사고 가해자의 연령대도 높아지고 있다. 일률적으로 운전면허 취득 후 10년이 지나면 적성검사를 받아야 하는 제도와, 매년 개정되는 도로교통법을 재 교육받을 기회가 없는 운전자들에게 교육의 기회를 제공할 수 있는 제도의 개선이 시급하다.

2. 시간대별 예측모델

2020년 시간대별 교통사고 발생 빈도수를 예측하기 위하여 2014~2018년 학습 데이터 셋으로 2019년 검증한 결과값의 최적화한 하이퍼-파라메터를 이용하여 〈Fig. 9〉와 같이 직관적으로 수용할 수 있는 예측 결과를 얻었다. 〈Fig. 9〉에서 볼 수 있듯이 출근시간대인 08~09시에서 퇴근시간대인 08~09시에서 교통사고 발생 꼭지점을 나타내고 있다.

출근시간대에는 2014년 11,569건, 2015년에는 12,113건으로 2016년에는 11,920건, 2017년에는 11,966건, 2018년에는 12,111건, 2019년에는 12,983건으로 증가하고 있으며, 2020년 예측에서는 12,564건으로 나타났다. 이전 시간대인 07~08시가 2014년에는 8,022건에서 2019년도의 7,826건으로 감소하고 2020년에는 7,576건으로 지속적으로 감소하고 있는 것으로 예측된 것과 비교하면, 08~09시에 매년 교통량이 증가한 것으로 분석된다.



<Fig. 9> forecasted of traffic accidents by time

출근시간 이후인 09시부터 18시까지 꾸준히 교통사고가 증가하고 있으며, 퇴근시간인 18~19시에 대폭으로 상승하는 경향을 나타내고 있다. 교통사고를 가장 많이 발생하는 시간대는 퇴근시간대인 18~19시이다. 2014년 15,694건이 발생하였고, 2015년에는 16,245건, 2016년에는 15,908건, 2017년에는 15,586건, 2018년에는 16,086건, 2019년에는 17,806건으로 증가하고 있으며, 2020년 예측에서는 18,753건으로 증가할 것으로 나타났다. 이러한 현상은 2018년부터 18~19시에 교통량이 많아졌다는 것으로 분석되며, 이는 일과 가정이 양립되는 사회학적 영향으로 퇴근시간이 빨라지고 있다는 것으로 해석된다.

매년 교통사고가 감소하는 시간대는 19시부터이다. 매년 19시 이후에는 교통사고 감소폭이 커지고 있으며 새벽 03시까지 줄어들고 있다. 이는 퇴근시간 이후 새벽시간대까지 사회적 활동이 줄어든 원인이라고 판단된다. 22시 이후에는 매년 2000건 이상의 교통사고 발생이 감소하는 것으로 나타났다.

3. 예측 모델 검증

본 연구과정에서 전 세계적으로 코로나-19사태로 사람의 이동과 경제활동이 멈추는 초유의 사태가 발생하였다. 우리나라에서도 예외가 아니어서 생활문화 패턴의 변화가 교통 환경의 변화를 가속화시키고 있으며, 특히 가정의 배달 수요가 이륜차 통행량과 배송업체 새벽 운행 증가로 이어지고 있으며, 출퇴근 직장인 또는 학부모들의 자가 차량 이용 경향도 뚜렷해지고 있다. 또한, 개인형 이동 장치 보급 확대와 공유업체를 통한 이용이 증가하여 새로운 교통안전 위협요인으로 부상하고 있다. 본 연구에서 2020년 교통사고 발생 빈도수는 224,986건으로 예측되었다. 하지만 교통 환경의 변화와 교통안전 정책의 강력한 추진으로 실제 교통사고 발생은 209,654건으로 예측치보다 6.8% 감소하였다.

국토교통부 자료에 의하면 전국 일평균 도로교통량이 8년 만에 감소세를 기록하여 15,187대로 전년대비 1.1%가량 줄어든 것으로 나타났다. 전국의 도로 교통량이 감소한 것은 2012년(-0.6%) 이후 처음이다. 세부적으로 살펴보면 승용차 교통량은 0.9% 줄었고, 화물차 교통량은 2.2% 증가하였다. 특히 버스 교통량은 전년보다 38.7% 줄어 가장 크게 감소하였다.

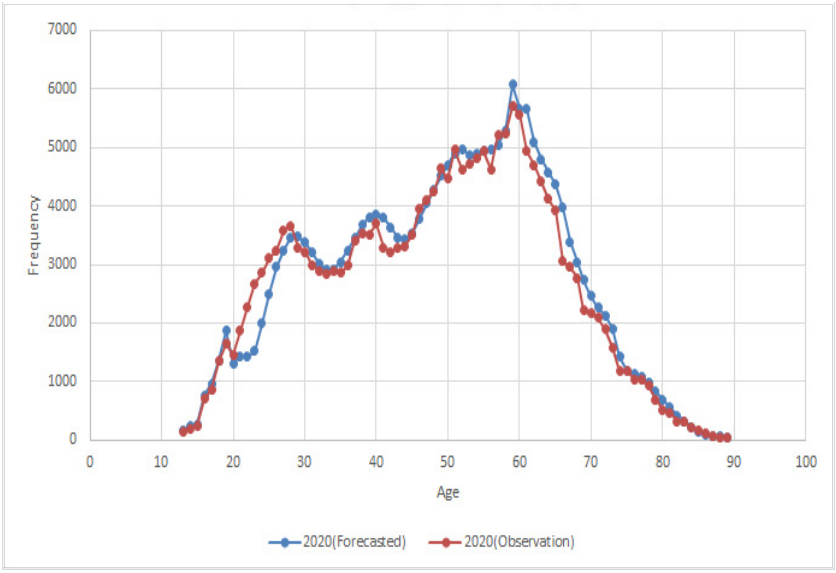
교통사고로 인한 피해도 줄어 2020년 코로나-19로 인하여 2019년에 비하여 교통사고 발생 빈도수는 8.7%가 감소

하였다. 이는 2014년부터 2019년까지 교통사고 발생 빈도수가 2.7% 증가하고 있는 패턴과 상반되게 진행되고 있어 LSTM에 의한 교통사고 발생 빈도수를 연령대별, 시간대별로 패턴의 변화를 검증하였다.

1) 연령대별 예측 검증

연령대별 예측에서 교통사고를 가장 많이 발생시키는 연령대는 2014년에는 53세가 6,581건, 2015년에는 54세가 6,941건, 2016년에는 55세가 6,416건, 2017년에는 56세가 6,103건, 2018년에는 58세가 6,000건, 2019년에는 58세가 6,412건으로 매년 1년씩 연령이 증가하고 있으며, 태어난 연도로 환산하면 1961년과 1962년에 태어난 연령에서 가장 교통사고 가해자가 많다는 것을 알 수 있다.

2020년 예측에서는 59세에서 6,074건이 발생할 것으로 나타났다. 2020년 실제 교통사고 발생 빈도수에서 가장 교통사고를 많이 발생시킨 연령은 59세로 5,709건이다. LSTM 예측값과 실제값과의 차이는 6.0%로 도로교통량 감소값과 비슷하게 나타나고 있다. 이는 교통사고 발생 빈도수는 연령대별로 일정한 패턴을 가지고 있다는 것을 알 수 있다. 또한 도로교통량의 변화에도 민감한 반응을 보인다고 할 수 있다.



<Fig. 10> Comparison of forecasted and actual values by age

〈Fig. 10〉에서 예측값과 실제값에서 10대에서는 가장 교통사고를 많이 발생시키는 연령이 19세다. 매년 19세 연령에서는 운전면허를 처음 취득하는 연령이기도 하고, 2020년에는 19세부터 교통사고 발생이 예측값보다 실제값이 높은 이유는 코로나-19로 인한 배달문화의 발달로 20대 초반의 초보 운전자들이 이륜차 운행에 따른 교통사고 발생의 영향이 크다. 20대의 교통사고 발생이 다른 연령대와 달리 예측값이 실제값을 상회하고 있는 것도 현 교통 환경의 영향이라 할 수 있다.

2) 시간대별 예측 검증

<Fig. 11> Comparison of forecasted and actual values by time

〈Fig. 11〉에서 보는 바와 같이 출근시간대와 퇴근시간대 교통사고 발생 빈도수는 줄어드는 것으로 나타났다. 출근시간대인 8시~9시의 교통사고 발생은 지속적으로 증가하여 2020년 예측치에서도 12,564건으로 나타났으나, 실제 교통사고는 10,613건으로 16.2% 대폭 감소하였다. 이는 출근시간대의 교통량이 재택근무 등으로 감소한 것과 관련이 있다. 퇴근시간대인 18시~19시에는 예측값이 18,753건, 실제값은 16,806건으로 10.4% 차이가 발생하였다. 이는 코로나-19로 인한 정통적인 출퇴근 시간 개념이 변하고 있음을 나타낸다고 할 수 있다.

V. 결 론

본 연구에서는 연령대별, 시간대별 요인에 따른 교통사고 발생 빈도수 예측을 다루었다. 두 요인에 의한 교통사고 발생 빈도수 통계 값이 주기성을 갖는 시계열 데이터로 표현된다는 특징에 착안하여 LSTM 적용이 제안되었다. LSTM은 주기성을 갖는 시계열 데이터의 예측에 널리 이용되는 인공지능 알고리즘의 하나이다. 2014~2020년 기간 동안에 한국 내에서 발생한 1,549,151건의 교통사고 기록이 LSTM 학습과 정확도 평가에 이용되었다. 개발된 LSTM 모델을 이용하여 2020년에 대한 연령대별, 시간대별 교통사고 빈도수 예측이 이루어졌다.

2020년 코로나-19로 인하여 사회적 거리두기, 재택근무의 확산, 배달문화의 정착 등 교통 환경의 변화가 8년 만에 도로교통량이 1.1% 감소하고 교통사고 발생 빈도수도 8.0% 감소하는 등 기존과 다른 패턴이 발생하였다. 또한, 다양한 교통사고 발생을 줄이기 위한 안전 정책으로 전반적인 교통사고 발생이 줄어들었다. 그럼에도 불구하고 전체적인 교통사고 발생 빈도수는 감소하였지만, 매년 연령대가 1~2년씩 높아지면서 특정 출생년도에서 교통사고 발생이 높은 것으로 나타났다. 2020년 현재 만 59세가 교통사고를 가장 많이 내고 있으며, 매년 이 연령대가 여전히 사회활동이 많고 앞으로도 계속적으로 교통사고를 발생시킬 가능성이 높은 것으로 분석되었다. 이 연령대에 대한 특별한 안전교육 대책이 수립되어야 할 필요성이 있다. 그리고 운전면허를 처음 취득하는 만 19세에서 교통사고가 많이 발생하는 것에 대한 대안은 운전면허 응시 전 안전교육의 내실화가 필요하고, 학교 교육에서 교통안전교육 의무화가 시행되어야 한다.

또한 시간대별 분석에서도 현재의 사회적 분위기에 의하여 교통사고 발생 시간대가 변화하는 것으로 분석되었다. 여전히 8시~9시 출근시간대에 교통사고 발생 빈도의 꼭지점이 나타나고 퇴근시간대인 18시~19시 교통사고 발생 빈도

수가 높지만 그 높이가 코로나로 인해 많이 줄어드는 현상이 나타나고 있다. 본 논문에서 교통사고 발생 빈도수로 사회적 현상의 한 단면을 분석할 수 있다고 판단된다.

본 연구로 교통사고는 사회적 변화에도 민감하지만, 패턴을 가지고 발생한다는 것을 증명하였다. 전체적으로 코로나-19로 인한 도로교통량의 감소가 있었지만, 연령대별, 시간대별 교통사고 발생 패턴은 큰 변화가 없었다는 것을 증명하였다. 따라서, 특정 연령대를 대상으로 교통사고 예방 교육을 실시하거나, 처음 운전면허를 취득하는 19세를 대상으로 하는 맞춤형 교통안전 추진 전략도 필요하다. 또한, 교통사고 발생도 특정 시간대에 집중하고 있으며, 특정 지점과 구간에서 발생하고 있으므로 집중 단속과 계도가 필요하다. 이는 딥러닝을 통하여 추적할 수 있으며, 교통단속 특정 시간대와 특정 구간, 지점에 집중한다면 교통사고로 인한 인명과 재산 피해를 최소화 시킬 수 있을 것이다.

향후 LSTM을 통한 교통사고 예측 모델에서 다양한 딥러닝 알고리즘을 적용할 필요성이 있다. 전통적인 회귀모형과 딥러닝 알고리즘을 상호 비교하여 과거의 패턴에서 정확한 미래를 예측할 수 있는 알고리즘 개발이 필요하다. 또한, 교통사고는 인적 요인이 대부분을 차지하며 이러한 인적 요인에 대한 다양한 분석과 빅데이터를 통하여 인적 요인이 교통사고에 영향을 주는 사회적 환경에 대한 연구가 추가되어 져야 할 것이다.

REFERENCES

Armstrong J. S.(1985), *Long-rang Forcasting: From Cristal Ball to Computer*, 2nd, ed, Wiley.

Bengio Y., Courville A. and Vincent P.(2013), "Representation Learning : A Review and New Perspectives," *IEEE Trans. PAMI, Special issue Learning Deep Architectures*, vol. 35, no. 8, pp.1798-1828.

Benoit F.(2019), *Road accidents in Switzerland forecasting – A brief comparison between Facebook Prophet and LSTM neural networks*, Towards Data Science.

Chris Olah, https://colah.github.io/posts/2015-08-Understanding-LSTMs/, 2021.05.21.

Christian S., Christian A. T. and Dumitru E.(2013), "Deep neural networks for object detection," *Advances in neural processing systems*.

Christopher O.(2015), *Understanding LSTM Networks*.

Chung J., Gulcehre C., Cho K. and Bengio Y.(2014), "Empirical evaluation of gated recurrent neural netwotks on sequence modeling," arXiv preprint arXiv:1412.3555.

Guangyuan P., 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.

Han J. S.(2019), *A Study on the Development of a Traffic Accident Risk Prediction Model Using Deep Learning Techniques*, Korea University, pp.12-17.

Hinton G. E., Osindero S. and Teh Y. W.(2006), "The A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp.1527-1554.

Hochreiter S. and Schmidhuber J.(1997), "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp.1735-1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Honglei Ren. et al.(2017), "A deep learning approach to the prediction of short-term traffic accident risk,"arXiv preprint arXiv:1710.09543.

Jakub Kvita, https://kvitajakub.github.io/2016/04/14/rnn-diagrams/, 2021.04.12.

Jang W. J.(2019), *Applications of deep learning methods in transportation research*, Koti, pp.13-21.

Jason B.(2020), *Deep Learning for Time Series Forecasting*, Machine Learning Mastery, pp.123-160.

Koroad(2020), Traffic accident statistical analysis, http://taas.koroad.or.kr

KOTI(2018), "Results of Investigation on Victims of Traffic Accidents and Improvement Plans for Victim Support System," *National Assembly Traffic Safety Forum*.

Krizhevsky A.(2009), "Convolutional Deep Belief Networks on CIFAR-10," *Unpublished Manuscript*, vol. 1, pp.1-9.

Lee K. and Roh J.(2015), "Development of Traffic Accident Prediction Model Using Probability Parameter - Targeted at 4 Intersections in the Metropolitan Area and Busan Metropolitan City," *Journal of the Korean ITS Society*, vol. 14, no. 5, pp.101-111.

Lee S., Kim J. and Kim T.(2003), "Development of urban road traffic accident prediction model in the planning stage according to road and traffic characteristics," *Journal of the Korean Traffic Association*, vol. 21, no. 4, pp.133-144.

Makridakis S., Spiliotis E. and Assimakopoulos V.(2018), "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, e0194889, https://doi.org/10.1371/journal.pone.0194889.

National Police Agency(2004), *A Study on the Revision of Traffic Accident Statistics*, National Police Agency, pp.39-69.

Nikhil B. and Nicholas L.(2017), *Fundamentals of Deep Learning*, O'reilly, pp.88-115.

Oh J., Yoon I. Hwang J., Choi J. and Han E.(2014), "A comparative study of the performance of the traffic accident prediction model at the intersection of the four districts using nonlinear regression analysis, artificial neural network, and structural equation," *Journal of the Korean Traffic Association*, vol. 32, no. 3, pp.266-279.

Reimers N. and Gurevych(2017), "Optimal hyper-parameters for deep lstm-networks for sequence labering tasks," arXiv preprint arXiv:1707.06799.

Ryu J. D.(2018), *Development of Expressway Traffic Accident Prediction Model Using Deep Learning*, AJOU University, vol. 17, no. 4, pp.14-25.

Sameen M. I. and Pradhan B.(2017), "Severity Prediction of Traffic Accidents with Recurrent Neural Networks," *Applied Sciences*, vol. 7, no. 6, 476.

Tomas M., Martin K., Lukas B., Jan H. C. and Sanjeev K.(2010), *Recurrent neural network based language model*, Interspeech, pp.1045-1048.

Witten I. H. et. al.(2011), *Data Mining : Practical Machine Learning Tools and Techniques*, Third Edition, Morgan Kaufimann, Boston.

Zhihao Z. et al.(2020), "Traffic Accident Prediction Based on LSTM-GBRT Model," *Journal of Control Science and Engineering*, vol. 2020, https://doi.org/10.1155/2020/4206919.