

시공간 특성 기반 LSTM-MLP 모델을 활용한 교통사고 위험도 예측 연구

정현진¹, 양지웅¹, 홍정희^{2*}

¹연세대학교 컴퓨터정보통신공학부, ²연세대학교 소프트웨어학부

Spatiotemporal Feature-based LSTM-MLP Model for Predicting Traffic Accident Severity

Hyeon-Jin Jung¹, Ji-Woong Yang¹, Ellen J. Hong^{2*}

¹Department of Computer & Telecommunications Engineering, Yonsei University

²Division of Software, Yonsei University

요약 급격한 도시화와 기술의 발전으로 자동차 보급 대수가 급증하면서 교통사고가 빈번하게 발생하고, 이로 인해 인적 피해와 경제적 손실이 증가하고 있다. 따라서 교통사고의 예방과 사고로 인한 피해를 최소화하기 위해 교통사고의 위험성을 예측할 수 있는 기술이 필요하다. 교통사고는 교통 혼잡도, 교통 환경, 도로 상태를 포함한 다양한 요인들로 인해 발생한다. 이러한 요인들을 기반으로 교통사고는 시공간적인 특성을 가지게 된다. 본 논문은 교통사고 데이터를 분석하여 교통사고의 주요 특성들을 파악하고, 이를 기반으로 시계열 형식의 데이터로 재구성하였다. 그리고 시공간적인 특성을 우수하게 파악할 수 있는 LSTM-MLP 기반 모델을 구축하여 교통사고 예측에 활용하였다. 실험을 통해 제안한 모델은 기존의 교통사고 위험도 예측 모델보다 더 합리적이고 정확한 예측함을 입증하였다. 본 논문에서 제안하는 교통사고 위험도 예측 모델은 내비게이션 등의 도로 상황과 환경을 실시간으로 파악할 수 있는 시스템에 적용할 수 있다. 이를 통해 도로 사용자들의 안전성을 향상시키고 교통사고로 인한 사회적 비용을 최소화하는 데 기여할 것으로 기대된다.

• 주제어 : 교통사고 예측, 시공간 특성, LSTM-MLP 모델, 실시간 예측, 교통 데이터

Abstract Rapid urbanization and advancements in technology have led to a surge in the number of automobiles, resulting in frequent traffic accidents, and consequently, an increase in human casualties and economic losses. Therefore, there is a need for technology that can predict the risk of traffic accidents to prevent them and minimize the damage caused by them. Traffic accidents occur due to various factors including traffic congestion, the traffic environment, and road conditions. These factors give traffic accidents spatiotemporal characteristics. This paper analyzes traffic accident data to understand the main characteristics of traffic accidents and reconstructs the data in a time series format. Additionally, an LSTM-MLP based model that excellently captures spatiotemporal characteristics was developed and utilized for traffic accident prediction. Experiments have proven that the proposed model is more rational and accurate in predicting the risk of traffic accidents compared to existing models. The traffic accident risk prediction model suggested in this paper can be applied to systems capable of real-time monitoring of road conditions and environments, such as navigation systems. It is expected to enhance the safety of road users and minimize the social costs associated with traffic accidents.

• Key Words : Traffic accident prediction, Spatiotemporal characteristics, LSTM-MLP model, Traffic Data

Received 23 November 2023, Revised 11 December 2023, Accepted 15 December 2023

* Corresponding Author Ellen. J. Hong, Division of Software, Yonsei University, 1 Yonseidae-gil, Wonju, Gangwon-do, Korea.
E-mail: ellenhong@yonsei.ac.kr

I. 서론

현대 사회에서의 급격한 도시화와 기술의 발전은 자동차 보급 대수의 급증에 크게 기여하고 있다. 이는 도로의 차량 흐름 복잡성을 상승시킬 뿐만 아니라 교통사고 발생률 상승세의 원인이 된다. 2018년 세계보건기구(WHO)에서 발행한 ‘2018 도로 안전 세계현황 보고서’ [1]에서는 전 세계의 교통사고 사망자가 135만 명을 넘었다고 밝혔다. 이처럼 교통사고 발생률이 증가함에 따라 막대한 인명 피해와 경제적 손실은 심각한 문제로 대두되고 있다.

교통사고 발생 위험 예측은 교통의 최적화, 안전한 경로 제공 및 효율적인 비용으로 교통 인프라를 개선을 위해서 지난 몇십 년 동안 많은 연구가 이루어졌다. 하지만 기존의 연구는 데이터의 특성을 고려하지 않았거나 1시간 단위로 발생할 교통사고를 예측하여 실시간 교통사고 발생 위험도를 예측하지 못한다. 교통사고 발생률은 시간과 날씨, 해당 시간대의 교통 상태와 주변 환경에 따라서 실시간으로 변화한다. 따라서 실시간 교통사고를 예측하기 위해서는 도로의 특성, 교통 제한 속도와 같은 정적인 데이터뿐만 아니라 날씨, 교통 속도, 태양 고도와 같이 실시간으로 변화하는 동적인 데이터의 특성을 고려해야만 한다.

본 논문은 2020년 1월부터 2021년 12월까지 총 2년간 서울특별시에서 발생한 교통사고 데이터(발생 일시, 위치)와 교통 속도, 날씨, 도로 길이, 차선 개수, 태양 고도, 태양 방위각, 과속방지턱, 단속 카메라와 같이 다양한 데이터를 수집하여 사용한다.

본 논문은 RNN 기반의 LSTM과 DNN 기반의 MLP를 융합한 모델을 통해 기존 연구들의 문제점을 해결하고자 한다. 본 논문에서 제안한 모델은 시간 및 도로 상황에 따라 달라지는 동적 feature와 도로의 정적 feature를 활용하여 5분 단위로 도로 별 교통사고 발생 위험도를 실시간으로 예측한다.

II. 관련연구

인공신경망(ANN)을 사용한 교통사고 위험 예측 모델이 전통적인 통계 기법 기반 교통사고 예측 모델보다 개선된 성능을 보여준다는 사례로 Kim et al.(2022)[2]이 존재한다. Kim et al[2]에서는 음이항 회귀모형, 포아송 회귀 모형의 통계 기반 모형이라는 한계를 극복하기 위해 DNN

기반의 교통사고 예측 모델을 개발하였다.

교통사고는 다양한 환경 요인 및 도로의 상황에 의해 영향을 끼칠 수 있다. Lee et al.(2015)[3]는 교통사고와 기상 요인이 밀접한 연관성이 있음을 증명하기 위해 logistic regression과 decision tree를 활용하여 연구하였다. 그리고 Park and Hong(2022)[4]은 서울시의 도로 정보 데이터와 기상 데이터 및 도로 주변 건물들의 영향 요인을 동적 데이터와 정적 데이터로 구분하여 MLP 모델을 활용하여 연구하였다. MLP는 McClelland et al.(1986)[5]에 의해 제시된 인공신경망이다. 기존의 퍼셉트론이 선형 분류기라는 한계에 의해서 XOR 문제를 해결하지 못하는 것을 해결하기 위해 고안된 이론으로 기존 퍼셉트론 구조에서 은닉층(Hidden Layer)을 중간에 추가한다. MLP의 구조는 Figure 1.를 통해 확인할 수 있다.

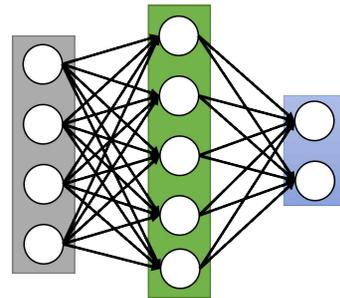


Fig. 1. MLP Structure

Noh et al.(2021)[6]는 LSTM 모델을 이용하여 교통사고 데이터를 연령대별, 시간대별로 나누어 교통사고 발생 빈도를 예측하는 연구를 진행하였다. LSTM은 Hochreiter et al.(1997)[7]에 의해서 증명된 RNN (Recurrent Neural Network)의 변형이다. LSTM은 기존 RNN의 장기적인 정보를 기억할 수 없다는 단점을 보완하기 위해서 시간 역전과 알고리즘을 통해 장/단기 정보를 기억할 수 있도록 설계한 신경망 구조이다. 표준 LSTM은 다음과 같이 표기된다. 각 시간 단계(step)를 나타내는 t 와 해당 입력 시퀀스는 $X = \{x_1, x_2, \dots, x_t\}$ 이다. 입력 게이트(input gate)는 i_t 이고, 삭제 게이트(forget gate)는 f_t , 출력 게이트(output gate)는 o_t , 메모리 셀의 상태는 c_t 이다. 해당 공식은 식 (1)과 같다.

$$\begin{aligned}
 i_t &= \sigma(W_i x_t + U_i h_t) \\
 f_t &= \sigma(W_f x_t + U_f h_t) \\
 o_t &= \sigma(W_o x_t + U_o h_t) \\
 \tilde{c}_t &= \tanh(W_c x_t + U_c h_t)
 \end{aligned} \tag{1}$$

단위 시간 t에서의 메모리 셀 상태 c_t 는 식 (2)와 같다.

$$c_t = i_t \circ \tilde{c}_t + f_t \circ c_{t-1} \tag{2}$$

메모리 셀의 상태가 변경된 후, 현재 은닉층 h_t 를 계산한다. 해당 계산은 식(3)을 사용한다.

$$h_t = o_t \circ \tanh(c_t) \tag{3}$$

여기서 W 는 입력의 가중치 행렬, U 는 상태 전이 가중치 행렬, σ 는 sigmoid 함수, \tanh 은 hyperbolic tangent 함수, h_t 는 출력의 상태 벡터, \tilde{c}_t 는 조정 및 변경 후의 새로운 셀의 상태, \circ 는 내적을 의미한다. 이러한 LSTM의 게이트 구조를 통해서 시계열 정보가 안정적인 장기적 의존성을 형성할 수 있게 된다. LSTM의 구조는 Figure 2.을 통해 확인할 수 있다.

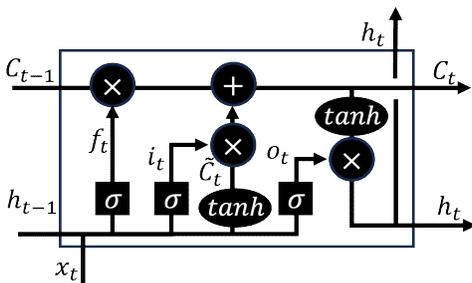


Fig. 2. LSTM Structure

그리고 Zhang et al.(2020)[8]에서는 LSTM-GBRT 기반 교통사고 예측 모델 구축 및 교통사고 데이터와 영향 요인 데이터들을 학습하여 도로 별 교통사고 안전 수준 지표를 예측하는 연구를 진행하였다.

많은 연구에서 딥러닝을 이용하여 교통사고 예측을 하였으며 대표적인 사례로 DNN 기반 모델 및 RNN 기반 모델이 있었다. 하지만 교통사고의 발생은 시간대

별로 영향이 미칠 수 있다는 점과 도로 상황 및 기상 요인 등 복합적인 요인들에 의해 영향을 미칠 수 있지만, 대부분의 연구는 이러한 요인들의 분석에 있어서 한계가 존재한다. 본 논문은 교통사고의 시간대별 교통 상황 및 환경 요인과 도로 특성 등을 포함한 복합적인 요인들의 학습 문제를 극복하고자 한다.

III. 교통사고 데이터

본 논문은 2020년 1월부터 2021년 12월까지 총 2년간의 서울특별시 교통사고 데이터를 바탕으로, 해당 기간의 정적 데이터인 도로 정보 및 도로 안전 시설물, POI와 동적 데이터인 환경 요소 및 교통 속도, 날씨를 포함한 데이터를 활용하여 연구를 수행한다. 데이터는 국가 기관의 다양한 포털 사이트로부터 수집되었고, 날씨 데이터의 경우 교통사고 발생과의 상관관계 분석을 통해 수정된 후 사용되었다. 본 논문에서 사용된 데이터는 Table. 1을 통해 확인할 수 있다.

Table 1. Dataset

Class	Factor	Feature	Description
Static	Road info	lanes	Number of lanes(1~7)
		speed_limit	Speed limit(km)
		length	Length of the road(m)
	Road safety facility	bump	Speed bump(0 or 1)
		camera	Speed camera(0 or 1)
	POIs	Description in Table 2	
Dynamic	Environment	sun_diff_angle	Sun's azimuth(°)
		sun_elevation	Sun's altitude(°)
	Traffic speed	speed	Current average speed on road(km/h)
	Weather	Description in Table 3	

3.1 Data Collection

도로와 관련된 데이터는 ITS 국가 교통정보 센터의 오픈 데이터 목록 중 표준 노드 링크[9]에서 제공하는 전국 단위 데이터를 서울특별시에 해당하는 표준 노드 링크를 활용하였고, 도로의 차선, 제한 속도 및 도로의 길이 데이터는 링크마다 주어진 특성들을 통해 수집할

Table 2. POI feature extraction (min : 0 ~ max : 10)

POIs Feature	Factor	POIs Feature	Factor
P1	Door-to-door sales business	P6	Snack bar
P2	Mail order business	P7	General restaurant
P3	Golf practice range business	P8	Bakery
P4	Fitness center business	P9	Clinic
P5	Group feeding facility	P10	Pharmacy

수 있었다.

교통사고 데이터는 ITS 국가교통정보센터의 오픈 데이터 목록 중 돌발 상황 정보[10]에서 일별로 제공하는 교통사고 데이터를 수집하여 월별 데이터로 재구성하였다. 비 교통사고 데이터의 경우, 2020년 1월부터 2021년 12월까지의 기간 중에서 교통사고가 발생하지 않은 시각과 서울특별시 범위 내의 표준 노드 링크를 임의로 추출하여 생성하였다. 과속방지턱 및 단속 카메라 데이터는 ‘LOCALDATA-지방행정인허가’의 생활 밀착 데이터[11]를 활용하여 도로의 정적 요소 특징으로 사용하였다.

POIs는 도로 인근의 병원 및 음식점과 같은 시설들의 집합이다. POIs의 특정 요소는 도로의 교통 복잡성을 증가시킬 수 있음에 따라 교통사고의 위험도에 영향을 미친다[12]. 이러한 이유에 근거하여 ‘LOCALDATA-지방행정인허가’의 POIs 데이터를 활용하였다. 하지만 제공하는 POIs 데이터는 필요 이상의 특징들을 포함하고 있다. 따라서, 빈도수 분석을 통해 도로의 혼잡성을 증가시킬 수 있는 특징을 추출하는 과정이 요구된다. 추출된 특징은 Table. 2에 기술하였다.

태양의 방위각 및 태양 고도 데이터는 각 교통사고와 비 교통사고의 링크와 시간을 활용하여 Equation of Time, Hour Angle 및 Solar Declination을 계산하여 수집하였다. 교통 속도 데이터는 ITS 국가 교통정보 센터의 오픈 데이터 목록 중 차량검지기정보[13]에서 2020년 1월부터 2021년 12월까지 총 2년의 전국 링크별 5분 단위 평균 속도 데이터셋을 활용하였다. 교통 속도 데이터 또한 표준 노드 링크에 대응될 수 있도록 매핑 작업을 수행하였다.

날씨 데이터는 기상자료 개방 포털에서 제공하는 종관기상관측(ASOS) 데이터[14]를 통해 수집하였다. 그러나, 총 26개의 feature로 구성된 날씨 데이터의 다양성으로 인해 교통사고와의 연관성이 있는 feature의 추출 작업이 필요하다. Lee et al.[15]의 Pearson 계수를

활용한 feature 추출을 위한 상관관계 분석 실험을 통해 Rain, Temperature를 포함한 총 8개의 feature를 추출할 수 있었다. 추출된 feature는 Table. 3에 명시되어 있다.

Table 3. Weather feature extraction

Extracted Weather Feature	Factor
Rain	Precipitation(mm)
Temp	Temperature(°C)
Humid	Humidity(%)
Visibility	Visibility(m)
Dew_point	Dew point(°C)
Cloud	Cloud cover(decile)
Vaper_press	Vaper pressure(hPa)
Ground_temp	Ground temperature(°C)

3.2 Normalization

각 feature의 원시 데이터는 최소-최대 정규화 (Min-Max normalization)를 통해 차원 간의 차이를 제거한다. 원시 데이터를 선형 변환하여 [0, 1] 구간으로 매핑하는데, 이를 위한 변환 공식은 식 (4)와 같다.

$$d = \frac{x - \min}{\max - \min} \quad (4)$$

여기서 max는 샘플 데이터에서 해당 feature의 최댓값을, min은 샘플 데이터에서 해당 feature의 최솟값을 나타낸다. x는 원시 데이터를 의미하며, d는 정규화된 데이터를 나타낸다.

IV. 교통사고 위험도 예측 모델

본 논문은 Hochreiter et al.[6]에 나타난 LSTM의 특성과 McClelland et al.[4]에 나타난 MLP의 특성을 각각 활용하여 교통사고 데이터의 동적 feature와 정적

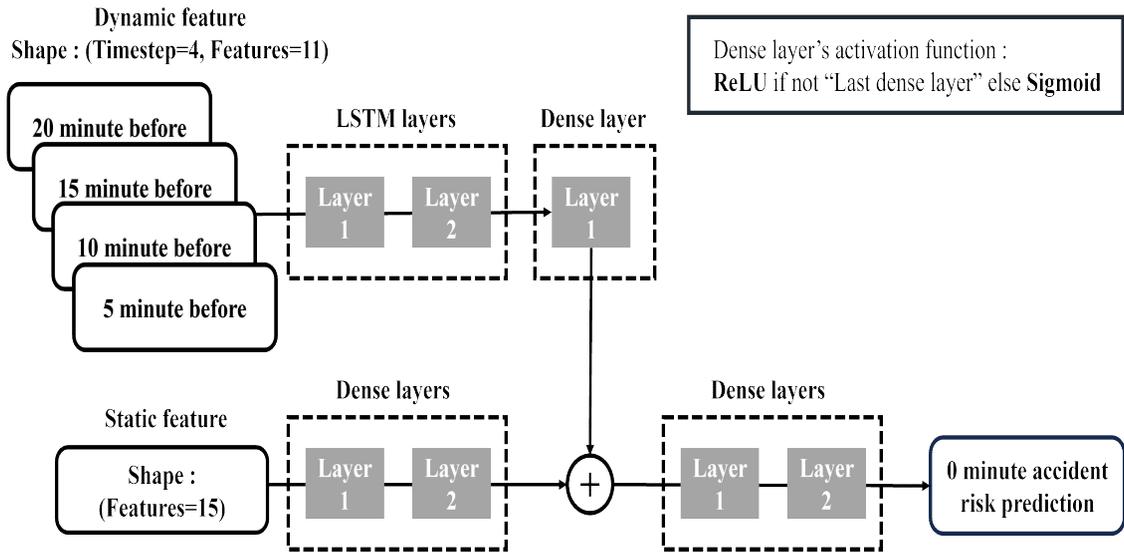


Fig. 3. Proposed Model Structure

feature를 모델 내에서 동시에 학습하는 것을 목표로 한다. 이러한 목표를 달성하기 위해 Figure. 3에 표현된 구조를 제안한다. 본 논문에서 제안한 모델은 LSTM과 MLP를 기반으로 구축되었다.

LSTM은 교통사고의 시간적인 변동성, 즉 동적 feature에 대한 학습이 효과적이다. 그리고, MLP는 비선형적인 관계를 가지는 교통사고의 정적 feature를 학습하는 데에 적합하다. 본 논문에서 제안한 모델은 이러한 각 모델의 장점을 활용하여 교통사고 데이터의 동적 feature 및 정적 feature를 효과적으로 학습하고 이를 종합하여 예측 결과를 도출하고자 한다.

교통사고의 동적 feature는 Table. 3의 날씨 데이터, 태양의 방위각 및 태양 고도 그리고 N분 전의 도로별 교통속도(N=5, 10, 15, 20)를 포함한다. “5, 10, 15, 20min before”는 교통사고가 발생 유무가 결정되기 5분 전, 10분 전, 15분 전, 20분 전을 의미한다.

교통사고의 정적 feature는 링크마다 주어진 특성인 차선 개수, 제한 속도, 도로의 길이를 포함할 뿐만 아니라 POIs, 과속방지턱 및 단속 카메라까지 포함한다. 정적 feature는 도로의 고정된 특성으로 시간에 따라 변하지 않는다. 따라서, 정적 feature는 시간대별로 구분되지 않고 일관되게 사용한다.

위와 같은 feature들을 기반으로, LSTM layer는 교통사고가 발생하기 20분 전, 15분 전, 10분 전, 5분 전의 동적 feature를 학습한다. 이후, Dense layer를 통해

LSTM layer의 출력값에 대한 학습을 진행하여 정적 feature와의 결합을 진행한다. 결합된 feature는 2개의 Dense layer를 통해 학습을 진행한 뒤 마지막 Dense layer를 거쳐 확률값을 출력한다. 확률값은 0분 전의 교통사고 발생 확률을 의미하며, 0.5 이상의 확률값은 교통사고가 발생할 것으로 예측한 경우이다. 반대로 0.5 미만의 확률값은 교통사고가 발생하지 않은 것으로 예측한 경우이다. 이러한 구조를 가지는 모델은 교통사고 위험도 예측을 위한 효과적인 도구로 활용될 수 있을 것이다.

V. 실험 및 분석

5.1 실험 설계

본 논문에서는 Keras에서 제공하는 LSTM과 MLP를 활용하여 제안하는 모델을 구현하였다. 동적 feature에 대한 LSTM layer와 Dense layer의 unit은 각각 128, 50으로 두었고, 정적 feature에 대한 Dense layer는 모두 50으로 설정하였다. 마지막 Dense layer는 0~1사이의 확률값을 출력하기 위해 Sigmoid 활성화 함수를 적용하였고, 그 외 Dense layer의 활성화 함수는 ReLU를 사용하였다. 손실함수는 ‘Binary Crossentropy’를 활용하였으며, 최적화 함수는 ‘Adam’으로 Learning Rate를 기본값으로 설정하였다.

Table 4. Evaluation Result

	Proposed Model	LSTM-MLP[14]	MLP[3]	LSTM
Loss	0.6702	0.6638	0.6916	0.7622
Accuracy	0.6001	0.5860	0.5907	0.5688
Recall	0.6302	0.6362	0.6570	0.5693
Precision	0.5944	0.5781	0.5801	0.5687
F1-score	0.6118	0.6058	0.6162	0.5690
Binary accuracy	0.6001	0.5860	0.5907	0.5688
AUC	0.6356	0.6337	0.6224	0.6045

본 논문에서 제안하는 모델과의 비교 실험을 위해 W.Zhu et al.(2021)[16]에서 제안한 LSTM-MLP 모델, Park and Hong(2022)[3]의 MLP 모델과 기본 LSTM 모델을 활용하였다.

데이터셋은 2020년 1월부터 2021년 12월까지 2년간의 서울특별시 교통사고 데이터 25,062개를 사용하였다. 또한, 교통사고 발생 시간을 기준으로 5분 전, 10분 전, 15분 전, 20분 전의 데이터셋을 생성하였다. 그리고 25,062개의 비 교통사고 데이터 또한 5분 전, 10분 전, 15분 전, 20분 전의 데이터셋을 생성하여 교통사고 데이터와 비 교통사고 데이터의 비율을 1:1로 맞추었다. 데이터셋은 Sklearn의 train_test_split을 활용하여 학습 데이터, 검증 데이터, 테스트 데이터로 나누었다. 이때 생성되는 데이터들의 비율을 8:1:1로 구성하였다.

실험은 Intel Core i9-12900KF 프로세서와 32GB의 RAM을 갖춘 Windows 10 운영체제가 설치된 컴퓨터에서 수행되었다. 그리고 TensorFlow 딥러닝 프레임워크를 활용하여 모델 학습과 평가를 수행하였다.

5.2 예측 모델 분석 결과

본 논문에서 제안하는 모델과 비교 모델의 성능을 비교하기 위해 Loss, Accuracy, Recall, Precision, F1-score, Binary accuracy, AUC를 평가지표로 활용한다. 실험 결과는 Table. 4에 기술되어 있다.

LSTM 모델은 비교 모델 중 모든 평가지표에 대해 가장 좋지 못한 성능을 보였다. 그에 비해, Park and Hong[3]의 MLP 모델은 Recall과 F1-score에서 가장 우수한 성능을 보였다. 그리고 W.Zhu et al.[14]의 LSTM-MLP 모델은 Loss에서 가장 좋은 성능을 보였으나 그 외의 평가지표에 대해서는 MLP에 비해 다소 낮은 성능을 기록하였다. 마지막으로, 본 논문에서 제안

한 모델은 Loss, Recall, F1-score을 제외한 평가지표들에 대해서 가장 우수한 결과를 보였다.

그리고 AUC-ROC Curve를 사용하여 모델의 AUC를 확인하고자 한다. ROC Curve는 분류기의 분류 능력을 그래프 좌표로 표현하는 방법이다. 실험 결과는 Figure. 4로 명시하였다. Figure. 4의 x축은 FPR(false positive rate)로 교통사고 데이터를 비 교통사고 데이터로 잘못 예측한 비율을 의미한다. 그리고 y축은 TPR(true positive rate)로 교통사고 데이터를 교통사고 데이터로 정확하게 예측한 비율을 의미한다.

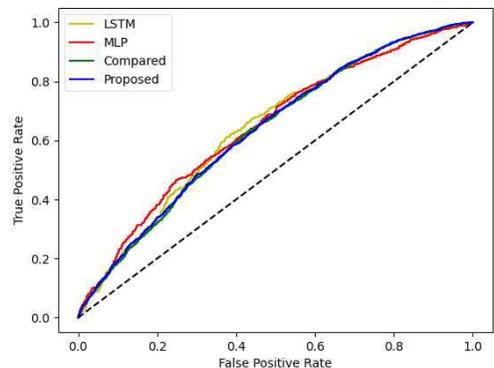


Fig. 4. ROC Curve Result

이때, 설정한 x축과 y축에 대한 ROC 곡선의 아래 면적은 AUC를 뜻한다. AUC가 1에 가까울수록 모델이 우수한 분류 성능을 나타낸다고 볼 수 있다.

Figure. 4에 따르면, 본 논문에서 제안한 모델은 비교 모델들보다 넓은 AUC 면적을 가지게 된다. 교통사고 예측 문제에 있어서 AUC는 가장 중요한 성능 지표로 활용됨에 따라 본 논문에서 제안한 모델이 가장 우수한 결과를 도출하였다.

VI. 결론

본 논문에서는 교통사고 문제의 해결을 위해 실시간 도로 상황과 데이터를 활용하여 교통사고 예측을 위한 모델을 제안한다. 기존의 교통사고 예측 모델은 시간대별 예측으로 실시간 예측에 제약이 있었으며, 다른 모델들은 전반적으로 교통 관련 데이터의 feature를 고려하는 데 있어서 한계점이 존재하였다.

본 논문은 기존 모델들의 한계점을 극복하기 위해 교통 관련 데이터를 동적 feature와 정적 feature로 구분하고, RNN 기반의 LSTM과 DNN 기반의 MLP를 융합하여 실시간 교통사고 발생 위험도를 예측하는 모델을 개발하였다.

본 논문에서 제안된 모델은 교통사고에 영향을 미치는 동적 요인이 분 단위로 주기성을 갖는 시계열 데이터로 표현되는 특성을 고려하여 LSTM을 적용하였으며, 정적 요인이 가지는 비선형적인 관계를 고려하여 MLP를 적용하였다.

제안된 모델은 AUC, Accuracy Precision, Binary accuracy 성능 지표를 기반으로 한 실험에서 우수한 결과를 보임으로써 5분 단위로의 도로 별 교통사고 위험도 예측 문제를 비교 모델들보다 잘 해결할 수 있음을 알 수 있었다.

본 연구의 결과를 토대로 LSTM과 MLP 기반의 제안된 모델은 도시 교통 안전성 향상을 위한 다양한 응용 분야에서 유용하게 활용될 수 있다. 예를 들어 교통 최적화, 안전한 경로 제공, 비용 효율적인 교통 인프라 구축 등에서 정책 수립 및 교통 인프라 개선에 필요한 의사 결정에 큰 도움을 줄 수 있을 것이다.

향후 연구 방향으로는 교통사고의 시공간적인 특징 뿐만 아니라 도로들의 인접한 도로들의 특징까지 파악하는 더 정확한 예측 모델의 개발이 필요하다. 특히 도심지에서 발생하는 교통사고는 인접한 도로들의 특징 파악이 더욱 중요할 수 있다.

따라서, 그래프 기반 모델을 활용하여 도로들의 인접한 도로들까지 파악할 수 있는 모델을 개발하여 교통사고에 영향을 미치는 요인들의 파악 범위를 확장하고자 한다. 또한, 분 단위보다 더 짧은 주기로 교통사고 위험 예측을 수행함으로써 더욱 정확한 실시간 교통사고 위험도 예측 모델을 개발하고자 한다.

ACKNOWLEDGMENTS

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1F1A1074273).

REFERENCES

- [1] World Health Organization, 2018, "Global status report on road safety 2018.", <https://www.who.int/publications/i/item/9789241565684>
- [2] U. Kim, Y. Kim, and J. Ko, "Development of Traffic Accident Prediction Model Based on Traffic Node and Link Using XGBoost," Journal of Society of Korea Industrial and Systems Engineering, vol. 45, no. 2, pp. 20-29, 2022.
- [3] K. Lee, I. Jung, Y. Noh, S. Yoon, and Y. Cho, "The effect of road weather factors on traffic accident - Focused on Busan area -," Journal of the Korean Data and Information Science Society, vol. 26, no. 3, pp. 661-668, 2015.
- [4] R. C. Park and E. J. Hong, "Urban traffic accident risk prediction for knowledge-based mobile multimedia service," Personal and Ubiquitous Computing, vol. 26, pp. 417-427, 2022.
- [5] J. L. McClelland, J. L., Rumelhart, D. E., & PDP Research Group, Parallel distributed processing, vol. 2, pp. 20-21, Cambridge, MA: MIT press. 1986.
- [6] 노유진, 배상훈, "LSTM을 이용한 교통사고 발생 패턴 예측," 한국 ITS 학회 논문지, vol. 20, no. 3, pp. 59-73, 2021.
- [7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
- [8] Z. Zhang, W. Yang and S. Wushour, "Traffic Accident Prediction Based on LSTM-GBRT Model," Journal of Control Science and Engineering, pp. 1-10, 2020.
- [9] ITS 국가교통정보센터, 2008.11, 전국표준노드링크, <https://www.its.go.kr/opendata/opendataList?service=none&elink>

- [10] ITS 국가교통정보센터, 2019.03, 공사·사고정보,
<https://www.its.go.kr/opendata/opendataList?service=event>
- [11] 지방행정인허가데이터:LOCALDATA, 과속방지턱 및 CCTV 데이터,
<https://www.localdata.go.kr/lif/lifeCtacDataView.do>
- [12] 조월, 이수기, “서울시 POI 빅데이터를 활용한 도시 활력과 영향요인 분석,” 국토계획, vol. 56, no. 7, pp. 87-102, 2021.
- [13] ITS 국가교통정보센터, 차량검지기정보, API
<https://www.its.go.kr/opendata/opendataList?service=detectorInfo>
- [14] 기상자료개방포털, 2015, 종관기상관측(ASOS),
<https://data.kma.go.kr/data/grnd/selectAsosRltnList.do>
- [15] J. Yang, H. Jeong, T. Kim, H. Lee and E. J. Hong, “Predicting traffic accident risk in Seoul metropolitan city : a dataset construction approach,” submitted.
- [16] W. Zhu, J. Wu, T. Fu, J. Wang, J. Zhang and Q. Shangguan, “Dynamic prediction of traffic incident duration on urban expressways: A deep learning approach based on LSTM and MLP,” Journal of intelligent and connected vehicles, vol. 4, no. 2, pp. 80-91, 2021.

저자소개

정 현 진 (Hyeon-Jin Jung)



2020년 3월~현재 : 연세대학교
컴퓨터정보통신공학부(재학)
관심분야 : 딥러닝, 머신러닝,
컴퓨터비전

양 지 웅 (Ji-Woong Yang)



2018년 3월~현재 : 연세대학교
컴퓨터정보통신공학부(재학)
관심분야 : 딥러닝, 머신러닝,
컴퓨터비전

홍 정 희 (Ellen J. Hong)



2013년 2월 : KAIST
전기및전자공학과(공학박사)
2016년 10월~2018년 1월 :
동서대학교 컴퓨터공학부 조교수
2018년 2월~2019년 8월 : KT
융합기술원 선임연구원
2019년 9월~현재 : 연세대학교
소프트웨어학부 조교수

관심분야 : 인공지능, 시스템 모델링 시뮬레이션,
시뮬레이션 기반 최적화, 디지털 트윈,
지능형 시스템