

## 날씨에 따른 교통사고 발생을 예측하는 Web Site 개발

조규철\*, 김 신<sup>o</sup>

\*인하공업전문대학 컴퓨터정보과,

<sup>o</sup>인하공업전문대학 컴퓨터정보과

e-mail: kccho@inhac.ac.kr\*, rlatks30@gmail.com<sup>o</sup>

### System Development of the Traffic Accident Prediction using Weather

Cho Kyu Cheol\*, San Kim<sup>o</sup>

\*Dept. of Computer Science, Inha Technical College,

<sup>o</sup>Dept. of Computer Science, Inha Technical College

#### ● 요약 ●

본 논문에서는 날씨와 상관관계를 갖는 교통사고에 대한 예측을 진행하는 Web Site 개발을 제안한다. 날씨에 영향을 받는 교통사고에 대한 일일 사망자 수, 교통사고 발생률의 각각의 예측값을 딥러닝 모델을 이용한다. 위의 모델을 작성하기 위하여 본 논문에서는 Anaconda 기반의 Jupyter Notebook에서 Python Tensorflow 모델을 작성하여 테스트하고, 만들어진 모델을 웹 사이트에서 불러오기 위해 Python 기반 Flask Web Framework를 통하여 웹 사이트를 개발한다. 개발된 웹 사이트는 사용자들은 Web Site에 날씨 정보를 입력하여 교통사고 발생률을 예측하고 볼 수 있다.

**키워드:** 정규화(Normalization), 다변량 회귀 모델(Multivariate Regression Model)

## I. 서론

날씨는 우리 사회에 영향을 끼치는 중요한 요소 중 하나이다. 또한, 교통사고 발생률과 교통사고로 인한 사망률의 경우도 날씨에 적지 않은 영향을 끼친다.

하지만 일반 사용자들이 쉽게 접근할 수 없는 영역의 데이터이다. 교통사고 발생률의 경우는 예측이 힘든 영역이기에 교통사고 발생률과 사망 사고 건수 등은 제공하지 않는다.

본 연구는 최저온도와 최고온도, 평균온도, 강수량과 같은 날씨 Data를 변인으로 가진 교통사고 발생률[1]에 대해 어떤 가중치를 갖는지, 실제 Data와의 예측한 Data의 차이는 얼마나 나는지 확인한다. 온도와 강수량에 따른 교통사고 사망률과 교통사고 발생률에 대한 자세한 상관관계와 모델을 TensorFlow[2]를 사용해서 작성하고, 이것을 기반으로 Flask Web Framework[3]로 사용자가 쉽게 접근할 수 있도록 Web Site를 개발한다.

## II. 날씨를 활용한 교통사고 예측 모델

### 1. Data-Set 준비

교통사고 발생률 예측 모델을 만들기 위해서는 우선 일 평균 교통사고 데이터[4]와 일별 날씨 데이터[5]가 필요하다. 이후 각각의 데이터를 모델에 사용하기에 최적화된 전처리 과정을 마친다. 본 논문은 한국의 2015년부터 2019년 12월까지 11,000개의 일별 데이터를 사용하였다.

### 2. 모델 개발

본 논문은 TensorFlow를 사용하여 모델을 개발하였다. 데이터를 불러오고 모델의 구성에 나서기 전 데이터 정규화[6]를 걸쳐서 피쳐들의 값을 일정한 범위로 맞춘다. 이를 통해서 일정하게 값을 조정하여 정확한 예측을 할 수 있다.

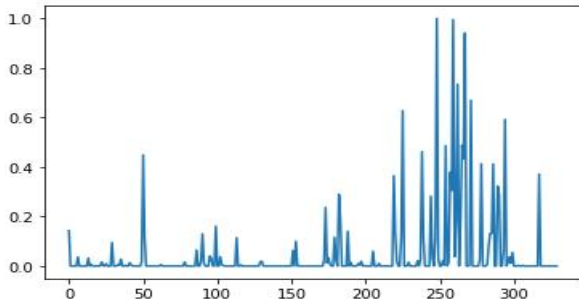


Fig. 1. 정규화를 걸친 데이터의 그래프

이후 다변량 회귀 모델(Multivariate Regression Model)을 사용한다. 단순 선형 회귀와는 다르게 종속변수가 두 개 이상일 때 다변량 선형 회귀 모델이라고 한다. Dense Layer를 층층이 쌓아 올려 모델의 구조를 완성한다.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 4)]	0
dense (Dense)	(None, 20)	100
dense_1 (Dense)	(None, 15)	315
dense_2 (Dense)	(None, 1)	16

Fig. 2. 완성된 모델의 구조

이후 실제 데이터를 이용해 모델을 학습한다. 이때 실제 값과 모델의 예측값의 오차를 줄이는 손실함수로는 MSLE를 사용한다. 위 모델의 최종 Loss(오차)는 500번의 학습으로 0.0761이다.

Table 1. Epoch and loss

Epoch	Loss
10	10.6857
50	5.1462
100	4.0942
150	2.3812
200	1.3650
300	0.9687
400	0.0897
500	0.0761

### 3. 예측값과 실제 값의 차이

본 논문에서 완성한 모델과 실제 값의 차이를 대표적인 시각화 도구인 matplotlib[7]를 사용하여 나타내고자 한다.

그림3과 같이 실제 값과 예측값이 유사한 흐름을 가진 것을 볼 수 있다. 이때, 실제 데이터에는 자동차 사용량이 급증하는 명절, 폭우로 인해 차량 운행이 적음 등의 아웃라이어 데이터가 포함되어있다.

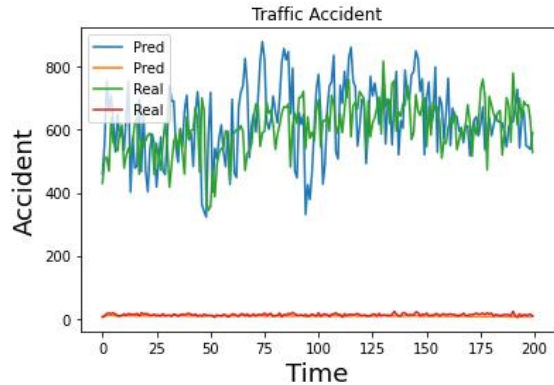


Fig. 3. 교통사고 예측값과 실제 값 비교 그래프

## III. 모델에 접근하는 웹 사이트

본 논문에서 Web-Site 구현은 Python을 기반으로 한 Flask Web Framework를 이용하여 개발하였다.

페이지는 사용자 편의성을 위해 1개의 페이지 내에서 진행되도록 하였으며, 메인 페이지 하단에는 입력과 출력을 나타내는 표현부가 따로 표출되어 있다.

날씨 데이터인 최저온도, 최고온도, 평균온도, 강수량을 입력하고 예측 버튼을 누르면 Html에서 Python으로 post 방식으로 값을 전달하기 때문에 서버의 Back-End인 Python에서 해당 값을 변수에 저장하고, 가상 개발 환경에서 저장된 모델을 불러와 예측한다. 예측한 값은 사용자에게 전달되기 전 보기 쉬운 형태로 가공되어 메인 페이지로 전송된다.

## IV. 결론

본 연구는 차후 모델의 예측 정확성을 높이기 위해서 혼동요인(confounding factor)을 통제하고 이를 위해서 설명변수를 추가적으로 모델에 추가하여 실제 값과의 오차를 더욱 줄일 수 있을 것으로 기대된다.

## REFERENCES

- [1] Ki-Kwang Lee. "Effects of Meteorological Factors on the Frequency of the Traffic Accidents in Seoul"
- [2] TensorFlow, <https://www.tensorflow.org/?hl=ko>
- [3] Flask Web Framework, <https://palletsprojects.com/p/flask/>
- [4] Traffic Accident Analysis System, <http://taas.koroad.or.kr/>
- [5] Weather Data Open Portal, <https://data.kma.go.kr/>
- [6] Huidea, , <https://huidea.tistory.com/39>
- [7] Mathlib, , <https://matplotlib.org/>