

RNN과 LSTM 기반의 PM10 예측 모델 성능 비교

정용진 · 이종성 · 오창현*

한국기술교육대학교

Performance Comparison of PM10 Prediction Models Based on RNN and LSTM

Yong-jin Jung · Jong-sung Lee · Chang-heon Oh*

Korea University of Technology and Education

E-mail : jungyj0211@koreatech.ac.kr

요 약

주관적 판단을 적용하여 예보되는 미세먼지 예보의 문제를 해결하기 위해 딥러닝 알고리즘을 이용하여 미세먼지 예측 모델을 설계하였다. 딥러닝 알고리즘 중 RNN과 LSTM을 이용하였으며, 하이퍼 파라미터 탐색을 통해 최적의 파라미터를 적용하여 설계하였다. RMSE와 예측 정확도를 통해 두 모델의 예측 성능을 평가하였다. 성능 평가 결과, RMSE와 전체 정확도에서 큰 차이는 없었으나 세부 예측 정확도의 차이가 있음을 확인하였다.

ABSTRACT

A particular matter prediction model was designed using a deep learning algorithm to solve the problem of particular matter forecast with subjective judgment applied. RNN and LSTM were used among deep learning algorithms, and it was designed by applying optimal parameters by proceeding with hyperparametric navigation. The predicted performance of the two models was evaluated through RMSE and predicted accuracy. The performance assessment confirmed that there was no significant difference between the RMSE and accuracy, but there was a difference in the detailed forecast accuracy.

키워드

RNN, LSTM, Deep learning, Particular matter, Prediction model

I. 서 론

심각한 미세먼지로 인해 다양한 사회적 문제가 야기됨에 따라 미세먼지를 사회적 재난으로 규정하여 미세먼지 저감 및 관리에 관한 특별법을 시행중이다[1]. 또한 미세먼지를 4단계로 구분한 AQI(air quality index)를 기준으로 정확하고 빠른 정보 전달을 위한 서비스를 제공하고 있다.

그러나 예측된 정보를 보완하기 위해 기상 예보관의 경험적 판단이 적용되며, 이에 따른 주관적 판단에 의한 오류 문제와 방대한 관측 데이터 검

토의 현실적인 문제가 존재한다[2].

본 논문에서는 이러한 문제를 해결하기 위해 미세먼지 농도의 예측을 위한 딥러닝 모델을 통해 성능 평가를 진행한다. 딥러닝 알고리즘 중 RNN(recurrent neural network)과 LSTM(long short term memory)을 이용하여 두 가지 예측 모델을 구축하고 두 예측 모델의 성능을 비교한다.

II. 데이터 구성 및 모델 설계

미세먼지 농도를 예측하기 위해 학습을 위한 데이터가 필요하며, 천안시에서 2009 ~ 2018년간 수

* corresponding author

집한 기상 데이터와 대기오염 물질 데이터 사용하였다[3][4]. 기상 데이터는 온도, 습도, 풍속, 풍향으로 구성하였으며, 대기오염 물질 데이터는 PM_{10} , O_3 , CO , NO_2 , SO_2 으로 구성하였다. 수집한 데이터들의 다양한 특성을 통일시키기 위해 one hot encoding과 min max scaling을 통해 전처리를 진행하였다. 구성된 데이터는 학습과 모델 성능 평가를 위한 75%의 훈련데이터와 25%의 실험데이터로 구성하였다. 최적화 함수는 adam, 활성화 함수는 ReLU를 사용하였다. 또한 학습 과정에서 과적합의 문제를 방지하기 위해 L2 규제와 dropout rate 등의 파라미터에 대한 최적화가 필요하며, 모델에 적용될 파라미터들과 하이퍼 파라미터 탐색을 진행하여 최적의 값을 도출하였다. 표 1은 RNN, LSTM 모델의 하이퍼 파라미터 탐색 결과이다.

표 1. 하이퍼 파라미터 탐색 결과

구 분	RNN	LSTM
layer	2	2
timesteps	24	24
hidden node	100	40
L2	0	0
dropout rate	0	0.2
batch size	20	40
epoch	100	100

III. 모델 성능 평가

구성한 학습 데이터와 파라미터 최적화가 진행된 RNN과 LSTM 기반의 예측 모델 구축 후 각 모델의 성능 평가를 진행하였다. 성능 평가는 RMSE(root mean square error)와 전체 정확도, AQI 수준별 정확도를 통해 진행하였다. 표 2는 RNN 모델과 LSTM 모델의 예측 성능 결과이다.

실제 PM_{10} 농도와 예측 농도 값의 평균 오차를 비교하기 위한 RMSE의 경우, LSTM 모델이 RNN 모델에 비해 약 $0.44\mu g/m^3$ 낮은 $10.2486\mu g/m^3$ 의 값을 보였다. 전체 예측 정확도의 경우, LSTM 모델이 RNN 모델에 비해 약 0.37% 높은 86.32%의 정확도를 보였다. AQI 수준별 정확도의 경우, ‘ 좋음’과 ‘매우 나쁨’의 경우 RNN 모델이 84.77%, 77.99%로 LSTM 모델보다 높은 정확도를 보였다. ‘보통’과 ‘나쁨’의 경우 LSTM 모델이 91.95%, 73.89%로 RNN 모델보다 높은 정확도를 보였다.

표 2. 예측 성능 평가

구 분	RNN	LSTM
RMSE	10.6854	10.2486
전체 정확도	85.95%	86.32%
‘ 좋음’ 정확도	84.77%	78.94%
‘보통’ 정확도	89.47%	91.95%
‘나쁨’ 정확도	72.10%	73.89%
‘매우 나쁨’ 정확도	77.99%	71.70%

IV. 결 론

본 논문에서는 딥러닝 알고리즘 중 RNN과 LSTM을 이용하여 두 가지 예측 모델을 구축하고 두 예측 모델의 성능 평가를 진행하였다. 예측 모델을 위한 데이터는 천안시에서 수집한 기상데이터와 대기오염물질 데이터를 사용하였다. 전처리를 통해 각 데이터에 대한 서로 다른 특성들을 동일한 특성으로 구성하였으며, 훈련데이터와 실험데이터를 각 75%, 25%로 구성하여 학습 및 실험을 진행 하였다. 각 모델에 대한 최적의 파라미터 값 도출을 위해 하이퍼 파라미터 탐색을 진행하였다. 모델의 성능 평가를 위해 RMSE, 전체 예측 정확도, AQI 수준별 정확도를 비교하여 진행하였다. 성능 평가 결과, 실제 값과 예측 값의 평균 오차의 경우 LSTM 모델이 약 $0.44\mu g/m^3$ 낮은 값을 보였다. 예측 정확도의 경우 RNN 모델이 ‘ 좋음’과 ‘매우 나쁨’의 수준에 대한 예측에서 LSTM 모델 보다 높은 정확도를 보였다. 각 모델 간의 RMSE와 전체 정확도에서 큰 차이는 없었으나 세부 예측 정확도의 차이를 보였다. 향후, 농도에 따른 특성 파악과 예측 모델의 적용을 통해 상대적으로 낮은 정확도를 가지는 ‘나쁨’과 ‘매우 나쁨’에 대한 예측 정확도 향상을 위한 연구를 진행할 계획이다.

Acknowledgement

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2019R111A3A01059038)

References

[1] 2020~2024 'Fine Dust Management Comprehensive Plan' Confirmed [Internet]. Available : <https://me.go.kr/cleanair/sub01.do>
 [2] Results of a specific audit by the Board of Audit

- and Inspection in 2017 [Internet]. Available :
http://www.kma.go.kr/info_open/cleango/gamsa01_list.jsp?bid=gamsa01&mode=view&num=41&page=1&field=&text=
- [3] Korea Meteorological Administration National Climate Data Center. Meteorological data open portal [Internet]. Available : <https://data.kma.go.kr>
- [4] Korea Environment Corporation. Air korea real time air quality [Internet]. Available : <https://www.airkorea.or.kr/realSearch>