딥러닝 알고리즘별 미세먼지 고농도 예측 성능 비교

이종성 · 정용진 · 오창헌*

한국기술교육대학교

Comparison of High Concentration Prediction Performance of Particulate Matter by Deep Learning Algorithm

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

Korea University of Technology and Education

E-mail : lee8611@koreatech.ac.kr

요 약

딥러닝을 이용하여 미세먼지 농도를 예측함에 있어 $81\mu g/m^3$ 이상의 고농도에 대한 특성이 예측 모델에 잘 반영되지 않는 문제가 있다. 본 논문에서는 딥러닝 알고리즘에 따라 고농도 영역에서의 미세먼지에 대한 특성 반영에 대한 결과를 확인하기 위해 예측 성능을 통한 비교를 진행하였다. 성능 평가 결과, 전반적으로 비슷한 수준의 결과를 보였으나, AQI 기준 “매우 나쁨”의 농도에서 RNN 모델이 다른 모델에 비해 보다 높은 정확도를 보였다. 이는 RNN 알고리즘이 DNN, LSTM 알고리즘보다 고농도에 대한 특성 반영이 잘 이루어진 결과를 확인하였다.

ABSTRACT

When predicting the concentration of fine dust using deep learning, there is a problem that the characteristics of a high concentration of $81\mu g/m^3$ or more are not well reflected in the prediction model. In this paper, a comparison through predictive performance was conducted to confirm the results of reflecting the characteristics of fine dust in the high concentration area according to the deep learning algorithm. As a result of performance evaluation, overall, similar levels of results were shown, but the RNN model showed higher accuracy than other models at concentrations of "very bad" based on AQI. This confirmed that the RNN algorithm reflected the characteristics of the high concentration better than the DNN and LSTM algorithms.

키워드

DNN, RNN, LSTM, Particular matter, Neural Network

Ⅰ. 서 론

미세먼지 농도 예측의 정확도를 향상시키기 위해 다양한 신경망 알고리즘을 이용한 연구들이 진행되고 있다. 그러나 다양한 환경 요인으로 인해 발생되는 미세먼지의 특성에 따라 $81\mu g/m^3$ 이상의 농도 값에 대한 예측 정확도가 $81\mu g/m^3$ 이하의 농도 예측의 정확도보다 낮은 수준을 보이는 문제가 있다[1][2]. 이는 $81\mu g/m^3$ 이상의 농도에 대한 특성이 예측 모델에 반영이 잘 되지 않는 것이라 볼 수 있다.

본 논문에서는 딥러닝 알고리즘을 이용하여 $81\mu g/m^3$ 이상의 고농도에 대한 특성 반영을 비교한다. 딥러닝 알고리즘 중 DNN(deep neural network), RNN(recurrent neural network), LSTM(long short-term memory) 알고리즘을 이용하여 $81\mu g/m^3$ 이상의 농도에 대한 예측 모델 설계를 통해 어떠한 알고리즘이 $81\mu g/m^3$ 이상 농도의 특성을 잘 반영하는지 성능 평가를 진행한다.

* corresponding author

II. 데이터 수집 및 전처리

각 모델의 학습을 위한 데이터는 기상 데이터와 대기오염물질 데이터를 사용하였다. 기상 및 대기오염물질 데이터는 천안시에서 2009년부터 2018년간 수집한 데이터를 사용하였다[3][4]. 기상 데이터 중 온도, 습도, 풍속, 풍향 데이터를 사용하였으며, 대기오염물질 데이터 중 PM_{10}, O_3, CO, NO_2, SO_2 데이터를 사용하였다. 이 데이터 중 $81\mu g/m^3$ 이상의 농도 값을 측정한 날짜의 데이터만 선별하였으며, one hot encoding과 min max scaling을 사용하여 전처리를 진행하였다. 데이터 전처리 후 모델의 학습과 성능 평가를 위한 데이터를 구분하기 위해 75%의 훈련데이터와 25%의 실험데이터로 구성하였다.

III. 모델 설계

딥러닝 알고리즘 중 DNN, RNN, LSTM 알고리즘을 이용하여 총 3개의 예측 모델을 설계하였다. 각 모델에 적용한 활성화 함수와 최적화 함수는 ReLU와 adam을 사용하였으며, RNN과 LSTM의 순회에 사용하기 위한 timesteps은 24로 설정하였다. 모델들의 최적의 성능을 위해 각각 하이퍼 파라미터 탐색을 통해 모델 세부 파라미터의 최적화를 진행하였다. 표 1은 하이퍼 파라미터 탐색 결과이며, 해당 결과를 반영하여 3가지 모델을 설계하였다.

표 1. 하이퍼 파라미터 탐색 결과

구 분	DNN	RNN	LSTM
layer	2	1	2
hidden node	100	80	80
L2	0.01	0	0
dropout rate	0.1	0	0.3
batch size	100	20	60
epoch	100	100	100

IV. 성능 평가

구성한 학습 데이터를 이용하여 예측 모델 학습 후 예측 정확도를 이용하여 성능 평가를 진행하였다. 표 2는 모델에 따른 고농도 예측 성능 결과이다. 전체 정확도와 AQI 기준으로 $81\mu g/m^3$ 이상에 해당하는 “나쁨”과 “매우 나쁨”의 정확도를 확인하였다. 각 모델별 전체 정확도는 LSTM 모델이 76.14%로 가장 높았으며 DNN 모델이 74.84%로 가장 낮은 정확도를 보였다. “나쁨”에 해당하는 농도의 정확도는 LSTM 모델이 76.83%로 가장 높았으며, RNN 모델이 75.24%로 가장 낮은 정확도를 보였다. “매우 나쁨”의 농도에서는 RNN 모델이 72.79%로 가장 높은 정확도를 보였으며, DNN 모델이 65.44%로 가장 낮은 정확도를 보였다. 전체 정확도의 경우 준수한 정확도는 아니지만 3가지 모델의 정확도가 비슷한 수준을 보였으나, “매우 나쁨”의 농도에서는 RNN 모델이 다른 모델에 비해 높은 성능을 보였다.

표 2. 모델별 고농도 예측 성능 결과

구 분	DNN	RNN	LSTM
전체 정확도 ($81\mu g/m^3$ ~)	74.84%	75.88%	76.14%
AQI “나쁨” (~ $150\mu g/m^3$)	75.50%	75.24%	76.83%
AQI “매우 나쁨” ($151\mu g/m^3$ ~)	65.44%	72.79%	66.18%

V. 결 론

본 논문에서는 미세먼지 $81\mu g/m^3$ 이상의 고농도에 대해 딥러닝 알고리즘별 예측 성능의 비교를 진행하였다. 딥러닝 알고리즘 중 DNN, RNN, LSTM 알고리즘을 이용하였으며, 모델의 학습을 위한 데이터는 천안시에서 10년간 수집한 기상 및 대기오염 물질 데이터를 활용하였다. 각 모델의 경우 최적의 성능을 위해 하이퍼 파라미터 탐색을 통해 설계하였다. 각 모델을 이용하여 미세먼지 고농도에 대한 예측 성능 평가 결과, 전반적으로 비슷한 수준의 결과를 보였으나, AQI 기준 “매우 나쁨”의 농도에서 RNN 모델이 다른 모델에 비해 보다 높은 정확도를 보였다. 이는 RNN 알고리즘이 DNN, LSTM 알고리즘보다 고농도에 대한 특성 반영이 잘 이루어진 결과로 볼 수 있다. 향후, 고농도 영역에서의 예측 성능을 통해 미세먼지 농도의 전 영역 정확도 향상에 대한 연구를 진행할 계획이다.

Acknowledgement

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2019R1I1A3A01059038)

References

[1] H. L. Kim, and T. H. Moon, “Machine Learning-based Fine Dust Prediction Model using Meteorological Data and Fine Dust Data, Weather forecast and earthquake notification system operation,” *Journal of the Korean Association of Geographic Information Studies*, Vol. 24, No. 1, pp. 92-111, Mar. 2021.

[2] K. H. Jeon, J. H. Lee, J. H. Park, H. J. Park, Y. H. Lee, M. S. Jung, H. S. Lee, K. P. Nam, J. S. Myoung, K. C. Choi, and T. H. Kim, “A Study of Data Accuracy Improvement for National Air Quality Forecasting (III),” *National Institute of Environmental Research*, Dec. 2016.

[3] Air Korea. Inquiry of final confirmed measurement data [Internet]. Available : https://www.airkorea.or.kr/web/pastSearch?pMENU_NO=123.

[4] Korea Meteorological Administation. Weather Data Opening Portal Synthetic Weather Observation [Internet]. Available : https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36.