EEMD 기반의 미세먼지 농도 예측 모델

정용진 · 이종성 · 오창헌*

한국기술교육대학교

Particular Matter Concentration Prediction Models Based on EEMD

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

Korea University of Technology and Education

E-mail : jungyj0211@koreatech.ac.kr

요 약

미세먼지 에측의 정확도 향상을 위해 다양한 연구가 이루어지고 있으나 미세먼지 농도에 따른 다양한 특성으로 인해 딥러닝 모델의 학습이 잘 이루이지지 않는 문제가 있다. 본 논문에서는 미세먼지 농도의 특성을 분해하여 특성을 반영하기 위한 EEMD 기반의 미세먼지 농도 예측 모델을 제안한다. 미세먼지 농도를 EEMD를 통해 분해 후, 각각 도출된 특성에 따른 예측 결과를 앙상블하여 최종 미세먼지 농도 값을 도출한다. 모델의 성능 평가 결과, 91.7%의 미세먼지 농도 예측 정확도를 확인하였다.

ABSTRACT

Various studies are being conducted to improve the accuracy of fine dust, but there is a problem that deep learning models are not well learned due to various characteristics according to the concentration of fine dust. This paper proposes an EEMD-based fine dust concentration prediction model to decompose the characteristics of fine dust concentration and reflect the characteristics. After decomposing the fine dust concentration through EEMD, the final fine dust concentration value is derived by ensemble of the prediction results according to the characteristics derived from each. As a result of the model's performance evaluation, 91.7% of the fine dust concentration prediction accuracy was confirmed.

키워드

Deep learning, Particular matter, Prediction model, DNN, EEMD

Ⅰ. 서 론

미세먼지로 인한 사회적 문제의 심각성이 대두됨에 따라 여러 정책과 예보 서비스가 시행되고 있다[1]. 이에 대중들에게 보다 정확한 예보 서비스 제공하기 위해 다양한 방법을 통한 연구가 진행되고 있다. 미세먼지 농도 예측의 정확도 향상을 위한 주요 연구 방법으로는 다양한 딥러닝 알고리즘을 이용한 연구가 수행되고 있다. 그러나 미세먼지의 다양한 특성으로 인해 정확도 향상의 어려움이 있다[2][3].

본 논문에서는 미세먼지 농도의 다양한 특성을 반영하여 예측하기 위한 EEMD(ensembel empirical mode decomposition) 기반의 미세먼지 농도 예측 모델을 설계하였다. 미세먼지 농도의 특성을 파악하기 위해 EEMD를 사용하여 n개의 IMF(intrinsic mode function)로 분해한 뒤, 분해된 IMF를 예측하는 n개의 모델을 앙상블하여 최종 미세먼지 농도 값 예측 및 평가를 진행한다.

Ⅱ. 데이터 구성

미세먼지 농도를 예측하기 위한 모델의 학습 데이터는 천안시에서 2009년도부터 10년간 수집한 기상 및 대기오염 물질 데이터 사용하였다[4][5]. 기상 데이터는 온도, 습도, 풍속, 풍향으로 구성하였고, 대기오염 물질 데이터는 PM_{10}, O_3, CO,

* corresponding author

NO_2, SO_2을 수집하여 구성하였다. 수집한 데이터를 예측 모델 학습에 용이하도록 전처리 과정이 필요하다. 따라서, one hot encoding과 min max scaling을 통해 전처리를 진행하였다. 전처리과정이 끝난 데이터를 이용하여 모델의 학습과 모델 성능 평가를 위해 전체 데이터 중 75%의 훈련데이터와 25%의 실험데이터로 나누어 구성하였다.

Ⅲ. EEMD 기반 예측 모델 설계

EEMD는 데이터의 비선형 복잡성을 효과적으로 포착할 수 있어 불규칙적으로 변화하는 미세먼지 농도의 변화 특성을 파악할 수 있다. 따라서 예측 목표인 PM_{10}을 EEMD를 통해 분해 및 모델 앙상블 과정을 거쳐 최종 농도값을 예측하기 위한 모델을 설계하였다.

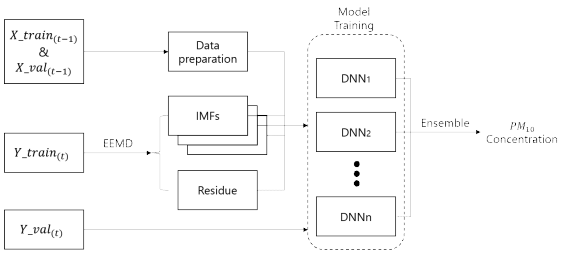

그림 1. EEMD 기반 예측 모델 구조

그림 1과 같이 훈련 데이터 중 PM_{10}을 EEMD 분해 결과에 따라 도출된 IMF 수만큼 DNN(deep neural network) 기반의 예측 모델을 생성한다. 생성된 DNN 모델들은 입력 데이터를 통해 해당하는 IMF를 예측한 후 각 모델에서 출력된 예측 값들을 앙상블하여 최종 PM_{10} 값의 예측을 진행한다.

표 1. 하이퍼 파라미터 탐색 결과

구 분	탐색 범위	탐색 결과
layer	2, 3, 4	3
hidden node	20 ~ 200	20
L2	0 ~ 0.001	0.001
dropout rate	0 ~ 0.5	0
batch size	20 ~ 100	60
epoch	100	100

각 DNN 모델은 회귀 모델이며 이에 따라 활성화 함수는 ReLu, 최적화 함수는 adam, 손실 함수는 mean square error를 사용하였다. DNN 모델의 최적화를 위해 세부 속성 값의 경우, 하이퍼 파라미터 탐색을 통해 공통으로 적용하였다. 표 1은 DNN 모델의 하이퍼 파라미터 탐색 결과이다.

Ⅳ. 성능 평가

표 2는 EEMD로 분해된 PM_{10}에 대한 n개의 IMF를 예측하는 n개 DNN 모델들의 앙상블로 PM_{10} 예측 결과이다.

표 2. 예측 성능 평가

EEMD 횟수	400	500
DNN 모델 수	16	15
RMSE	20.2329	18.4161
전체 정확도	89.41%	91.7%
‘좋음’ 정확도	88.12%	81.08%
‘보통’ 정확도	89.39%	94.22%
‘나쁨’ 정확도	92.77%	87.74%
‘매우 나쁨’ 정확도	66.67%	93.71%

예측 결과, 400회의 EEMD를 진행한 모델의 경우 농도 예측의 전체 정확도는 89.41%를 보였다. AQI 기준으로 구분하여 정확도를 분석한 결과, ‘나쁨’의 정확도가 92.77%로 가장 높았으며, ‘매우 나쁨’의 정확도가 66.67%로 가장 낮은 것을 확인하였다. 500회의 EEMD를 진행한 모델의 경우 91.7%의 전체 정확도를 보였으며, ‘보통’의 정확도가 94.22%로 가장 높았고, ‘좋음’의 정확도가 81.08%로 가장 낮은 것을 확인하였다. 두 경우를 비교하였을 때, ‘매우 나쁨’의 정확도에서 27.04%의 가장 큰 차이를 보였다.

Ⅴ. 결 론

본 논문에서는 EEMD 기반의 미세먼지 예측 모델을 구축하고 예측 성능 평가를 진행하였다. 예측 모델을 위한 데이터는 천안시에서 수집한 기상 및 대기오염물질 데이터를 수집하여 사용하였다. 수집한 데이터의 전처리 과정과 예측 모델의 학습을 위한 훈련데이터와 실험데이터를 각 75%, 25%로 구분하여 최종 데이터를 구성하였다. EEMD를 이용하여 훈련 데이터 중 PM_{10}의 분해를 진행하였으며, 분해 결과로 도출된 n개의 IMF를 예측하기 위한 DNN 모델을 구성하였다. DNN 모델들의 결과를 앙상블하여 최종 미세먼지 농도 예측 값을 도출하였다. EEMD 진행 횟수를 기준으로 400회와 500회의 예측 결과 비교를 통해 성능을 평가하였

다. 성능 평가 결과, 90% 수준의 전체 정확도를 보였다. AQI 기준으로 구분하여 정확도를 확인한 결과, '좋음'과 '나쁨'에 대한 정확도는 400회의 EEMD를 진행한 모델이 500회의 EEMD를 진행한 모델보다 높게 나왔으며, '보통'과 '매우 나쁨'에 대한 정확도는 500회의 EEMD를 진행한 모델이 더 높은 정확도를 보였다. 특히 '매우 나쁨'의 경우, 66.67%와 93.71%의 수치로 27.04%의 차이를 보였다. 결과를 통해 EEMD를 이용할 경우, 미세먼지 농도 범위에 따른 최적의 정확도를 도출할 수 있는 가능성을 확인할 수 있었다. 향후, 각 농도 범위에 따른 최적의 모델을 구성하여 미세먼지 농도 예측의 전체 정확도를 향상시키기 위한 연구를 진행할 계획이다.

Acknowledgement

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

References

[1] 2020~2024 'Fine Dust Management Comprehensive Plan' Confirmed [Internet]. Available : https://me.go.kr/cleanair/sub01.do

[2] H. Luo, D. Wang, C. Yue, Y. Liu, and H. Guo, "Research and Application of a Novel Hybrid Decomposition-ensemble Learning Paradigm with Error Correction for Daily PM10 Forecasting," *Atomospheric Research*, Vol. 201, pp. 34-45, 2018.

[3] Y. Bai, B. Zeng, C. Li, and J. Zhang, "An Ensemble Long Short-term Memory Neural Network for Hourly PM2.5 Concentration Forecasting," *Chemosphere*, Vol. 222, pp. 286-294, 2019.

[4] Korea Meteorological Administration National Climate Data Center. Meteorological data open portal [Internet]. Available : https://data.kma.go.kr

[5] Korea Environment Corporation. Air korea real time air quality [Internet]. Available : https://www.airkorea.or.kr/realSearch