

인공 신경망을 이용한 미세먼지 예측 모델

정용진¹ · 조경우¹ · 강철규² · 오창현^{1*}

¹한국기술교육대학교(KOREATECH) · ²SEMES

Particulate Matter Prediction Model using Artificial Neural Network

Yong-jin Jung¹ · Kyoung-woo Cho¹ · Chul-gyu Kang² · Chang-heon Oh^{1*}

¹Korea University of Technology and Education · ²SEMES Co. LTD.

E-mail : jungyj0211@koreatech.ac.kr

요 약

미세먼지에 대한 이슈가 확산됨에 따라 미세먼지에 대한 정보를 실시간으로 제공하는 서비스가 많아지고 있다. 그러나 미세먼지를 수집하기 위한 센서 노드의 결함이 발생할 경우 해당하는 서비스를 제공하지 못하는 경우가 발생한다. 이러한 문제를 해결하기 위해서는 미세먼지의 예측과 추론이 필요하다. 본 논문에서는 미세먼지를 예측하기 위해 과거의 미세먼지 및 기상 데이터를 기반으로 신경망 알고리즘을 이용하여 미세먼지 예측 모델을 설계한다. 또한 설계된 모델의 입력 데이터를 다양화하여 학습함으로써 예측 결과를 비교한다.

ABSTRACT

As the issue of particulate matter spreads, services for providing particulate matter information in real time are increasing. However, when a sensor node for collecting particulate matter is defective, a corresponding service may not be provided. To solve these problems, it is necessary to predict and deduce particulate matter. In this paper, a particulate matter prediction model is designed using artificial neural network algorithm based on past particulate matter and meteorological data to predict particulate matter. Also, the prediction results are compared by learning the input data of the model in the design stage.

키워드

Particulate matter, Deep learning, Artificial neural network, Prediction model

I. 서 론

미세먼지에 대한 사회적 이슈가 확산됨에 따라 미세먼지에 대한 정보를 제공하는 서비스가 많아지고 있다. 국내 기상청, 한국 환경 공단에서는 실시간으로 미세먼지에 대한 정보를 제공하고 있으며, 해당 기관에서 제공하는 데이터를 기반으로 여러 종류의 미세먼지관련 스마트 디바이스 어플리케이션이 등장하고 있다[1-3]. 그러나 미세먼지 데이터를 수집하기 위한 과정에서 측정 센서 및 네트워크망의 결함 발생 시 해당 서비스를 원활히

제공하지 못하는 문제가 발생한다. 이러한 문제를 해결하기 위해 손실된 미세먼지 데이터의 예측이 필요하며 이에 따른 예측 모델이 필요하다.

따라서 본 논문에서는 미세먼지를 예측하기 위해 인공 신경망을 활용하여 예측 모델을 설계한다. 이를 위해 미세먼지에 영향을 미치는 기상 데이터와 미세먼지 농도를 학습 데이터로 활용한다. 또한 예측 지역 인근 측정소의 미세먼지 농도 데이터를 활용, 입력 데이터를 다양화하여 학습함으로써 예측 결과를 비교한다.

* Corresponding author

II. 인공 신경망 모델 설계

인공 신경망이란 인간의 뉴런 구조를 모티브로 만든 딥러닝 알고리즘으로서 입력 값이 어떤 클래스에 속하는지를 구분하는 분류 ANN과 입력 정보로 다른 값을 예측하는 회귀 ANN으로 구분된다 [4]. 인공 신경망의 구조는 입력 계층, 은닉 계층, 출력 계층으로 구성되며, 은닉 계층을 한 개 이상 포함할 수 있다. 본 논문에서는 미세먼지에 대한 예측 모델을 설계하기 위해 회귀 ANN을 이용한다.

인공 신경망을 이용하여 미세먼지를 예측하기 위해 선행된 여러 연구에서는 공통적으로 온도, 습도, 풍향, 풍속의 기상 데이터를 필수 입력 데이터로 사용하고 있다 [5-9]. 이러한 연구 결과를 바탕으로 학습 데이터의 경우 천안시 병천면의 기상 데이터, 미세먼지 농도의 1년간의 과거 데이터를 사용하였다. 또한 미세먼지 농도 입력 변수의 개수에 따른 예측 성능을 평가하기 위해 약 20km 반경 측정소(성황, 백석, 오창) 세 곳의 미세먼지 측정값을 추가하여 총 8,404건의 학습 데이터를 사용하였다. 표 1은 학습에 사용된 데이터 셋의 예시이다.

표 1. 학습 데이터 셋 예시

기상 데이터				미세먼지 농도			
온도	습도	풍향	풍속	병천	성황	백석	오창
23.7	50	70	2.9	24	32	21	26

미세먼지 예측을 위한 인공 신경망 모델의 경우 1개의 은닉 계층으로 설계하였으며 최적화 함수로는 Adam gradient, 활성화 함수는 ReLU를 사용하였다. 또한 학습 데이터의 정규화를 위해 MinMaxScaler를 사용하였다. 학습 횟수의 경우 전체 데이터를 100개씩 나누어 총 100회의 학습을 진행하였으며, 예측 지역의 미세먼지 농도만을 학습한 경우와 인근 지역의 미세먼지 농도를 함께 학습한 경우와의 예측 성능을 비교하였다.

III. 분석 및 평가

그림 1의 a는 예측지역의 미세먼지 농도만을 학습하여 구축한 모델을 이용한 예측 결과이다. 실제 데이터와 비교하였을 경우 예측된 값이 변화의 추세는 반영하고 있으나 pm10 농도의 경우 30~50 ug/m³ 사이의 구간에서 표현되고 있다.

그림 1의 b는 a의 예측 모델 학습 시 인근 지역의 미세먼지 농도를 함께 학습한 모델을 이용한 예측 결과이다. 실제 데이터와 비교하였을 경우 전반적으로 비슷한 예측을 보여주고 있다.

그림 a와 b를 통해 미세먼지 예측 모델을 비교

한 결과 모델 학습 과정에서 기상 데이터와 예측 지역의 미세먼지 농도만 사용하여 진행할 경우 예측 정확성의 한계를 보여주고 있는 것을 확인하였다. 기상 데이터, 예측 지역의 미세먼지 농도 데이터와 함께 인근 지역의 미세먼지 농도 데이터를 추가하여 학습한 결과 전반적으로 예측의 정확도가 상승한 것을 확인할 수 있다.

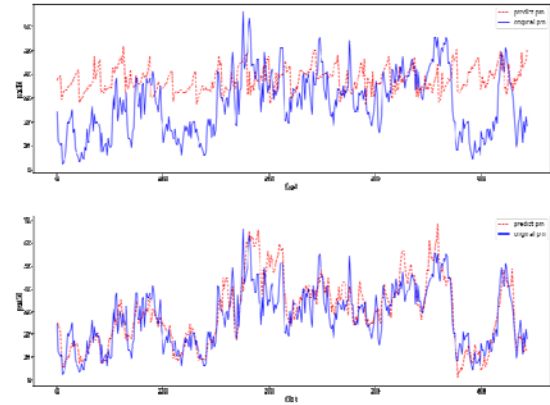


그림 1. 입력 데이터에 따른 미세먼지 예측 결과

IV. 결론

미세먼지로 인한 인체 영향으로 미세먼지 모니터링에 대한 중요성이 높아지고 있다. 그러나 미세먼지 측정 센서의 결함이나 네트워크망의 결함이 발생할 경우 미세먼지관련 서비스를 원활히 제공하지 못하는 문제가 발생한다. 이러한 문제를 해결하기 위해 손실된 미세먼지 데이터의 예측이 필요하며 이에 따른 예측 모델이 필요하다.

본 논문에서는 미세먼지를 예측하기 위해 인공 신경망을 활용, 기상데이터와 예측 지역의 미세먼지 농도만을 학습한 모델과 인근 지역 미세먼지 농도를 추가로 함께 학습한 두 가지 모델을 설계하여 예측을 수행하였다. 예측 결과 예측 지역의 미세먼지 농도만을 학습한 모델의 경우 예측 정확성의 한계를 보여주었으며, 인근 지역의 미세먼지 농도를 함께 학습한 모델의 경우 전반적으로 예측의 정확도가 상승한 것을 확인할 수 있었다.

향후 최적화된 학습 데이터를 이용하여 다양한 모델 설계를 진행함으로써 미세먼지 예측을 위한 최적의 모델 설계를 진행할 계획이다.

References

- [1] Korea Meteorological Administration National Climate Data Center. Meteorological data open portal [Internet]. Available : <https://data.kma.go.kr>
- [2] Korea Environment Corporation. Air korea real time

- air quality [Internet]. Available :
<https://www.airkorea.or.kr/realSearch>
- [3] Google Play Store [Internet]. Available :
<https://play.google.com/store/apps>
- [4] M. V. Gerven and S. Bohte, *Artificial Neural Networks as Models of Neural Information Processing*, Frontiers Media SA, 2018.
- [5] S. C. Park, M. J. Kim, M. H. Kim, H. G. Namgung, K. T. Kim, K. H. Cho and S. B. Kwon, "Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN)," *Journal of hazardous materials*, Vol. 341, pp. 75-82, Jan. 2018.
- [6] A. Alimissis, K. Philippopoulos, C. G. Tzani and D. Deligiorgi, "Spatial estimation of urban air pollution with the use of artificial neural network models," *Atmospheric Environment*, Vol. 191, pp. 205-213, Oct. 2018.
- [7] G. Grivas and A. Chaloulakou, "Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece," *Atmospheric Environment*, Vol. 40, No. 7, pp. 1216-1229, Mar. 2006.
- [8] F. Franceschi, M. Cobo and M. Figueredo, "Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering," *Atmospheric Pollution Research*, Vol. 9, No. 5, pp. 912-922, Sep. 2018.
- [9] J. W. Cha and J. Y. Kim, "Development of data mining algorithm for implementation of fine dust numerical prediction model," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 22, No. 4, pp. 595-601, Apr. 2018.