

다층 퍼셉트론 신경망을 이용한 미세먼지 예측

조경우¹ · 정용진¹ · 강철규² · 오창현^{1*}

¹한국기술교육대학교(KOREATECH) · ²SEMES

Particulate Matter Prediction using Multi-Layer Perceptron Network

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

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

E-mail : pinokio622@koreatech.ac.kr

요 약

미세먼지에 대한 인체 영향에 관한 사회적 관심이 높아짐에 따라 미세먼지 예측 알고리즘의 필요성이 증가되었다. 많은 연구에서 기상 데이터를 이용하여 통계 모델링 및 기계 학습 기법 기반 예측 모델이 제안되었으나, 해당 모델의 환경 및 세부조건을 정확히 설정하기는 어렵다. 또한 국내 기상 측정소 데이터의 경우 누락된 데이터가 존재하여 새로운 예측 모델을 설계해야 할 필요성이 있다. 본 논문에서는 미세먼지 예측을 위한 선행 연구로서 다층 퍼셉트론 신경망을 활용하여 미세먼지 예측을 수행한다. 이를 위해 측정소 3곳의 기상 데이터를 기반으로 예측 모델을 설계, 실제 데이터와의 비교를 통해 미세먼지 예측을 위한 알고리즘의 적합성을 평가한다.

ABSTRACT

The need for particulate matter prediction algorithms has increased as social interest in the effects of human on particulate matter increased. Many studies have proposed statistical modelling and machine learning techniques based prediction models using weather data, but it is difficult to accurately set the environment and detailed conditions of the models. In addition, there is a need to design a new prediction model for missing data in domestic weather monitoring station. In this paper, fine dust prediction is performed using multi-layer perceptron network as a previous study for particulate matter prediction. For this purpose, a prediction model is designed based on weather data of three monitoring station and the suitability of the algorithm for particulate matter prediction is evaluated through comparison with actual data.

키워드

Particulate matter, Multi-layer perceptron, Neural network, Deep learning

1. 서 론

공장, 자동차, 일상생활 등에서 사용되는 화석연료로 인해 발생하는 많은 배출 가스는 미세먼지 발생의 주요 원인으로 지적되고 있다. 또한 미세먼지의 인체 영향으로 인해 사회적 관심이 높아지고 있으며, 이로 인해 미세먼지 예측 알고리즘의 필요성을 증가시키고 있다 [1],[2].

미세먼지 예측의 경우 많은 연구에서 기상데이터를 활용한 통계 모델링 기법과 기계 학습 방법을 사용하여 미세먼지 농도 예측 및 분석이 진행되었다 [3-5]. 그러나 해당 모델을 활용하기 위해 측정 환경 및 세부조건을 정확히 설정하는 데는 어려움이 따른다 [6]. 또한 국내 측정소에서 제공하는 대기 환경 데이터의 경우 시스템의 유지 보수, 고장 등으로 인해 누락된 데이터가 존재하여 기존 연구의 결과를 활용하기 어려운 문제가 있어 새로운 예측 모델을 설계해야 할 필요성이 있다. 이에

* Corresponding author

본 논문에서는 미세먼지 예측을 위한 선행 연구로서 다수의 연구에서 활용된 다층 퍼셉트론 신경망을 활용하여 미세먼지 예측을 수행한다. 이를 위해 측정소 3곳의 기상 데이터를 기반으로 예측 모델을 설계, 실제 데이터와의 비교를 통해 미세먼지 예측을 위한 알고리즘의 적합성을 평가한다.

II. 다층 퍼셉트론 신경망 모델

다층 퍼셉트론 신경망은 input layer와 out layer 사이에 하나 이상의 hidden layer가 존재하는 계층 구조의 신경망이다. 네트워크는 input, hidden, out 방향으로 연결되어 있으며, 각 layer내의 연결과 out layer에서 input layer로의 직접적인 연결은 존재하지 않는 전방향 네트워크이다.

본 논문에서 사용된 다층 퍼셉트론 신경망 모델은 그림 1과 같다. Input layer의 입력 데이터의 경우 학습에 사용될 4개의 기상 데이터와 3개 측정소의 미세먼지 농도 데이터, 측정 시간으로 총 8개의 데이터를 활용하였다. Hidden layer의 경우 8, 10, 8의 뉴런을 가지는 3개의 layer로 설계되었으며, 활성화 함수는 ReLU, 최적화 함수는 adam을 사용하여 학습을 수행하였다. 이후 out layer를 통해 예측된 미세먼지 농도를 출력한다.

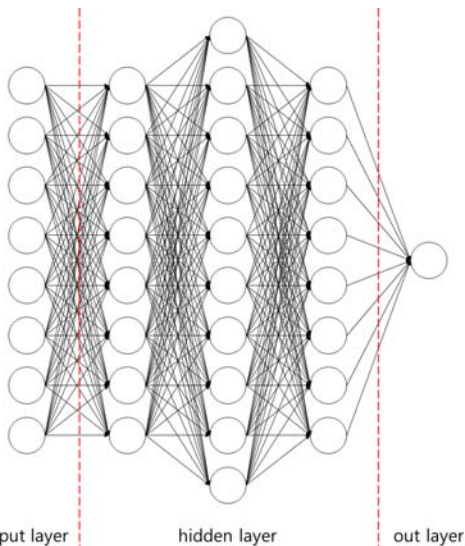


그림 1. MLP 기반 미세먼지 예측 모델 구조

III. 실험 및 평가

미세먼지 측정 데이터의 경우 국내외에서 대기 환경 모니터링 시스템을 구축하여 측정 정보를 제공하고 있으며, 국내의 경우 전국 367개의 측정소에서 측정된 미세먼지 데이터를 제공하고 있다 [7,8]. 본 논문에서는 천안 지역의 기상 데이터 중 시간대 별 온도, 습도, 풍향, 풍속 데이터와 한국기

술교육대학교 1캠퍼스 근방 미세먼지 측정소인 오창, 성황, 백석 측정소의 미세먼지 농도 및 측정시간을 입력 변수로 사용하였다. 이때, 누락된 데이터만을 제거한 뒤 추가 전처리 없이 사용하였다. 학습은 전체 데이터를 100개씩 나누어 학습하도록 하였으며, 총 100회의 학습을 진행하였다.

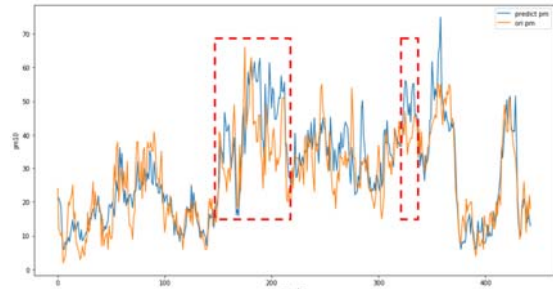


그림 2. 미세먼지 예측 결과

그림 2는 미세먼지 예측 결과와 실제 미세먼지 농도와의 비교 그래프이다. 예측 결과를 확인했을 때, 특정 부근의 경우 실제 농도 값과 상승, 하강 추세가 반대로 나타나는 문제가 발생했으며, 미세먼지 농도 오차의 경우 평균 $6.6\mu\text{g}/\text{m}^3$, 최대 $32.4\mu\text{g}/\text{m}^3$, 최소 $0.0001\mu\text{g}/\text{m}^3$ 을 나타내었다. 그러나 전체적인 예측 추세는 비슷한 모습을 나타내었다. 또한 상승, 하강 추세가 반대로 나타난 영역의 경우 약 $30\sim 75\mu\text{g}/\text{m}^3$ 의 고농도에 해당하는 특징을 보였다. 이는 미세먼지 데이터의 비선형적 특성에 따른 문제로, 고농도의 미세먼지에 대한 상관성을 고려한 전처리가 이루어져야 함을 의미한다. 따라서 고농도의 미세먼지에 대한 기상 데이터 간의 상관성 분석, 데이터 전처리, 추가 알고리즘과의 앙상블을 통해 다층 퍼셉트론 신경망을 사용할 경우 전체적인 예측 성능을 높일 수 있을 것으로 판단된다.

IV. 결론

미세먼지의 인체 영향으로 인해 미세먼지 예측 알고리즘의 필요성이 높아지고 있다. 다수의 연구에서 기상 데이터를 활용하여 통계 모델링, 기계 학습 기법 기반 예측 모델을 설계하였으나, 모델에 맞는 측정 환경 및 세부조건을 정확히 설정하기 어렵고, 국내의 경우 측정소에서 제공하는 데이터가 누락되는 문제가 있어 새로운 예측 모델을 설계해야 할 필요성이 있다. 본 논문에서는 미세먼지 예측을 위한 선행 연구로서 다층 퍼셉트론 신경망을 활용한 미세먼지 예측을 수행하였다. 예측 결과, 전체적 예측 추세는 비슷한 경향을 보였으나 고농도의 미세먼지 영역에서는 상승, 하강 추세가 반대로 나타나는 문제가 발생하였다. 따라서 고농도의 미세먼지에 대한 기상 데이터간의 상관성 분

석 등의 전처리 후 다층 퍼셉트론 신경망을 이용할 경우 전체적인 예측 성능을 높일 수 있을 것으로 기대된다.

References

- [1] Ministry of Environment, *Comprehensive measures for particulate matter control*, 2017. 9.
- [2] Y. K. Jang, "Status and problems of particulate matter pollution," *Journal of Environmental Studies*, vol. 58, pp. 4-13, 2016. 9.
- [3] A. Chaloulakou, G. Grivas, and N. Spyrellis, "Neural network and multiple regression models for PM10 prediction in athens: A comparative assessment," *Journal of the Air & Waste Management Association*, vol. 53, no. 10, pp. 1183-1190, 2003.
- [4] S. Munir, M. T. Habeebullah, A. R. Seroji, E. A. Morsy, A. M. Mohammed, W. A. Saud, A. E. A. Abdou, and A. H. Awad, "Modeling particulate matter concentrations in Makkah, applying a statistical modeling approach," *Aerosol and Air Quality Research*, vol. 13, no. 3, pp. 901-910, 2013.
- [5] M. M. Dedovic, S. Avadakovic, I. Turkovic, N. Dautbasic, and T. Konjic, "Forecasting PM10 concentrations using neural networks and system for improving air quality," in *Proceedings of 2016 XI International Symposium on Telecommunications (BIHTEL)*, Sarajevo, pp. 1-6, 2016.
- [6] K. T. Sohn, and D. H. Kim, "Development of statistical forecast model for PM10 concentration over Seoul," *Journal of the Korean Data And Information Science Society*, vol. 26, no. 2, pp. 289-299, 2015. 04.
- [7] C. L. Muller, L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R.R. Leigh, "Crowdsourcing for climate and atmospheric sciences: current status and future potential," *International Journal of Climatology*, vol. 35, no. 11, pp. 3185-3203, 2015.
- [8] Air Korea – Station information [Internet]. Available : <https://www.airkorea.or.kr/stationInfo>