
저자 (Authors)	이민우, 최형림, 하정호, 이승호, 안병구 Minwoo Lee, Hyunglim Choi, Jeongho Ha, Seungho Lee, Beongku An
출처 (Source)	한국통신학회 학술대회논문집 , 2019.11, 495-496(2 pages) Proceedings of Symposium of the Korean Institute of communications and Information Sciences , 2019.11, 495-496(2 pages)
발행처 (Publisher)	한국통신학회 Korea Institute Of Communication Sciences
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09277796
APA Style	이민우, 최형림, 하정호, 이승호, 안병구 (2019). 머신러닝을 활용한 미세먼지 예측. 한국통신학회 학술대회논문집, 495-496
이용정보 (Accessed)	부산도서관 210.103.83.*** 2021/09/24 13:53 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

머신러닝을 활용한 미세먼지 예측

이민우, 최형림, 하정호, 이승호, 안병구*

홍익대학교

*beongku@hongik.ac.kr

Fin Dust Prediction Using Machine Learning

Minwoo Lee, Hyunglim Choi, Jeongho Ha, Seungho Lee, Beongku An*

Hongik University

요약

매년 심각해지는 미세먼지를 ANN 인공지능망 알고리즘을 적용한 머신러닝을 이용하여 예측하는 방법을 제안한다. 제안된 방법의 특징과 기여도는 다음과 같다. 첫째, 기상정보를 데이터화 하여 데이터셋 확보를 한다. 둘째, 아두이노를 통해 현재의 미세먼지 농도를 데이터화 한다. 그리고 미세먼지에 영향을 끼치는 요인들을 적용하여 만들어진 ANN알고리즘으로 현재 측정 데이터를 가지고 미래의 미세먼지를 예측한다. 제안된 방법으로 미래의 미세먼지 농도를 측정한 값과 예측한 값을 비교함으로써 머신러닝이 얼마나 정확하게 미세먼지를 예측하는지 알 수 있다.

I. 서론

먼지란 대기 중에 떠다니거나 흩날려 내려오는 입자상 물질을 말하는데, 석탄, 석유 등의 화석연료를 태울 때 공장, 자동차 등의 배출가스에서 많이 발생한다. 먼지는 입자의 크기에 따라 $50\mu\text{m}$ 이하인 총먼지(TSP, Total Suspended Particles)와 입자크기가 매우 작은 미세먼지(PM, Particulate Matter)로 구분한다. 미세먼지는 다시 지름이 $10\mu\text{m}$ 보다 작은 미세먼지(PM10)와 지름이 $2.5\mu\text{m}$ 보다 작은 초미세먼지(PM2.5)로 나뉜다. PM10이 사람의 머리카락 지름($50\sim 70\mu\text{m}$)보다 약 $1/5\sim 1/7$ 정도로 작은 크기라면, PM2.5는 머리카락의 약 $1/20\sim 1/30$ 에 불과할 정도로 매우 작다.

미세먼지가 우리에게 미치는 영향은 상당히 거대하다. 미세먼지가 우리 몸속으로 들어오면 면역을 담당하는 세포가 먼지를 제거하여 우리 몸을 지키도록 작용하게 되는데, 이때 부작용인 염증반응이 나타난다. 기도, 폐, 심혈관, 뇌 등 우리몸의 각 기관에서 이러한 염증반응이 발생하면 천식, 호흡기, 심혈관계 질환 등이 유발 될 수 있다. 특히 노인이나 유아, 임산부에게는 일반인보다 더 많은 영향을 받는다. 그래서 인간은 나날이 심각해지는 미세먼지의 영향을 최대한 피하려고 노력하여야 한다.

미세먼지 예측을 위한 많은 연구는 지금도 진행되고 있다. 미세먼지 예측에 사용되는 대표적인 알고리즘은 신경망 알고리즘이다. 신경망을 이루고 있는 계층의 구성과 학습에 사용하는 변수와 방법에 따라 상이하다. 3-Layer(input, hidden, output)구성과 기상요인을 이용하여 PM10을 예측[1], 분류나 회귀분석에 사용되는 랜덤포레스트 모델을 이용한 대기질 분석[2], 각 대기질을 DNN(Deep Neural Network)으로 학습시켜 PM10, PM2.5를 예측하기도 하였다[3].

본 논문은 ANN 알고리즘을 이용하여 공개된 기상요인을 통해 PM10을 예측하기 위해 시스템을 구현하였다.

II. 본론

그림 1은 본 논문에서 제안하고 있는 ANN알고리즘을 이용한 미세먼지 예측 시스템의 구성도이다.



그림 1. 시스템 구성도

그림 1에서와 같이 Anaconda에서 제공하는 패키지를 이용하여 ANN 알고리즘을 구현 후 확보한 Dataset을 통해 이 알고리즘을 학습시킨다. 본 논문에서 구현한 ANN 알고리즘은 그림 2와 같이 3층의 Neural Network로 구성하였다.

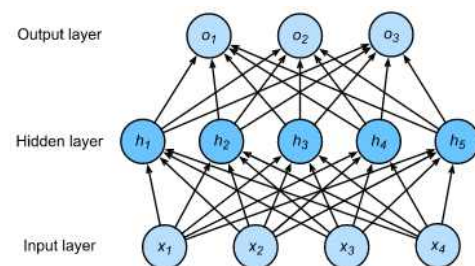


그림 2. Neural Network 구조

즉 기상요인을 반영하는 입력계층, 반복되는 학습을 통해 가중치로 패턴을 파악하는 은닉계층, PM10 수치를 예측하는 출력계층으로 구성되었다. 학습에 사용된 기상요인은 풍향, 풍속, 기온, O3, SO2, CO, SO4, PM2.5 총 8개 항목이다. 학습이 완료된 알고리즘과 현재 기상상황 및 Arduino 미세먼지 키트(GP2Y1014AU)로 실측한 값을 이용하여 예측 후 반복학습 시킨다.

III. 성능 평가

본 장에서는 II장에서 제안한 ANN 알고리즘을 이용한 미세먼지 예측 시스템을 구현하고 그 성능평가를 실시한다. 확보한 Dataset을 Training_set과 Test_set으로 나눈 후 구현이 완료된 ANN 알고리즘을 Training_set을 통해 학습시킨다. 그림 3은 학습이 완료된 알고리즘으로 Test_set을 예측할 때 손실 값과 Training_set을 통해 알고리즘을 학습시킬 때 매 Epoch(전체 데이터가 한 번 학습되는 단위)마다 손실 값 그래프를 나타내고 있다.

Test Loss -> 9.54

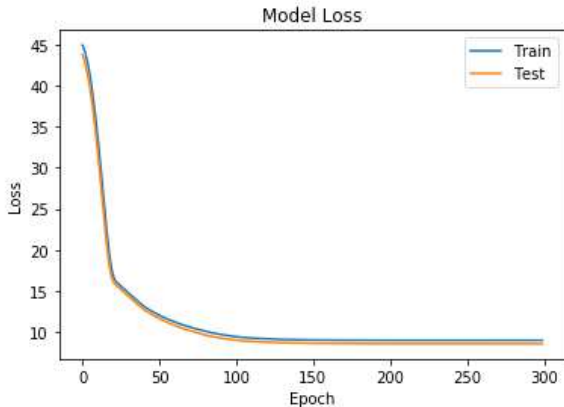


그림 3. 학습된 알고리즘의 손실 값 및 학습 중 알고리즘의 손실 값 그래프

그림 3의 그래프에서 볼 수 있듯 학습이 진행됨에 따라 손실 값이 낮아지는 것을 볼 수 있다. 또한 학습된 알고리즘의 예측 손실 값(Test Loss)도 확인할 수 있다. 하지만 손실함수로 사용한 평균 절대오차 MAE(Mean Absolute Error)의 값이 8~9 정도로 머무는 것을 보아 실제 값과 예측 값의 PM10 수치차이가 평균적으로 해당 값 정도인 것을 알 수 있다.

IV. 결론

본 논문에서는 머신러닝 알고리즘을 여러 요인들로 학습시켜 미세먼지 농도(PM10)을 예측하는 시스템을 제안하였다. 미세먼지 예측에 주로 사용되는 공개된 기상요인들(풍향, 풍속, O3, SO2 등)을 통하여 알고리즘을 반복 학습시킴으로써 손실을 최소화하여 정밀한 예측이 가능토록 하였다. 본 시스템은 ANN 알고리즘만을 사용하여 구현하였지만, 추후 분류 알고리즘을 추가, 더 많은 Dataset 확보, 더 적절한 Neural Network 구성, 하이퍼 파라미터 방법(모델을 구성하는 여러 요소 중 최적의 요소 값을 찾아내는 과정)을 통해서 설정해주어야 할 더 적절한 값을 찾아내는 등 추가적인 방법으로 알고리즘을 더 정밀하게 학습시켜 예측할 때 더 낮은 손실을 기대해볼 수 있다.

ACKNOWLEDGMENT

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF), Ministry of Education, under Grant (No. 2016R1D1A1B03934898) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(No. 2019R1A2C1083996).

참 고 문 헌

- [1]Kyunghak Cho, Byoung-young Lee, Myeongheum Kwon, Seogcheol Kim, "Air Quality Prediction Using a Deep Neural Network Mode," Journal of Korean Society for Atmospheric Environment Vol. 35, No.2, April 2019.
- [2]Jinwook Cha, Jangyoung Kim, "Development of Data Mining Algorithm for Implemetation of Fine Dust Numerical Prediction Model," Journal of the Korea Institute of Information and Communication Engineering Vol. 22, No.4, Aprill 2018.
- [3]강은하, "ANN(Artificial Neural Network)를 사용한 수원시 미세먼지 농도 예측," 수원시정연구원, October 2017.