

# 순환 신경망을 이용한 미세먼지 AQI 지수 예측

정용진 · 이종성 · 오창현\*

한국기술교육대학교

## Prediction of Particulate Matter AQI using Recurrent Neural Networks

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

Korea University of Technology and Education

E-mail : jungyj0211@koreatech.ac.kr

### 요 약

미세먼지에 따른 행동 지침을 위해 AQI 지수가 개발되어 사용되고 있다. AQI 지수에 대한 정보는 일반인들도 쉽게 제공 받을 수 있으며, 이에 따라 AQI 지수를 기반으로 다양한 서비스가 제공되고 있다. 서비스가 제공됨에 따라 정확한 AQI 지수의 예측이 필요하다. 본 논문에서는 미세먼지의 AQI 지수를 예측하기 위해 순환 신경망을 이용하여 분류 모델의 설계를 진행한다. 설계된 모델의 평가를 위해 실제 미세먼지와 예측치의 AQI 지수를 비교한다.

### ABSTRACT

The AQI index has been developed and used to guide the action of particulate matter. Information on the AQI index can be easily provided to the general public, and various services are provided based on the AQI index. As services are provided, accurate AQI index prediction is needed. In this paper, we design the classification model using the circular neural network to predict the AQI index of particulate matter. For the evaluation of the designed model, compare the AQI index of the actual particulate matter with the predicted value.

### 키워드

Recurrent Neural Network, particulate Matter, Neural Network, AQI

## I. 서 론

대기질과 사회 활동의 관계에 대한 선행연구에 따르면 대기질은 건강에 영향을 미치는 중요한 요인이며 이에 따라 사회 활동을 감소시키는 요인으로 분석하고 있다[1][2]. 최근 급증한 미세먼지로 인해 호흡기 질환, 심혈관 질환 등 건강에 대한 문제와 도시 경제 및 사회 활력 저하의 원인으로 심각한 사회적 이슈로 대두되고 있다[3]. 이러한 선행 연구들과 사회적 이슈에 따라 미세먼지에 따른 행동 지침을 위해 AQI(Air Quality Index) 지수가 개발되어 사용되고 있다[4]. AQI 지수에 대한 정보는 일반인들도 쉽게 제공 받을 수 있으며, AQI 지수를 기반으로 다양한 서비스가 제공 되고

있다. 이러한 서비스가 제공됨에 따라 정확한 AQI 지수의 예측이 필요하다.

본 논문에서는 인공신경망 알고리즘 중 지도학습기반의 순환신경망을 사용하여 AQI 지수를 예측하는 예측 모델을 설계한다[5]. 이를 위해 다수의 기상인자와 미세먼지 농도값을 이용하여 예측 모델을 설계하고 예측 모델의 실험 및 평가를 진행한다.

## II. 예측 모델 설계

예측 모델 설계를 위해 순환 신경망에 대한 지도학습이 필요하며, 학습을 위한 데이터가 필요하다. 사용된 데이터는 천안 지역의 2009년 9월부터 2018년 9월에 해당하는 과거 10년간 일평균 기상

\* corresponding author

데이터, 대기오염 물질 데이터를 활용하였다. 기상 데이터는 온도, 습도, 평균 풍속, 최대 풍향으로 구성되며, 대기오염 물질 데이터는  $O_3$ ,  $NO_2$ ,  $CO$ ,  $SO_2$ ,  $PM_{10}$  등의 물질로 구성하였다. 구성된 데이터를 이용하여 지도학습을 위해 AQI 지수를 종속 변수로 설정하였으며, 그 외의 나머지 데이터는 독립변수로 설정하여 사용하였다. 독립변수 중 최대 풍향의 경우 16방위의 범주형 데이터로서 하나의 벡터형 정수로 표현하기 위해 one-hot-encoding을 진행하였으며, 각각의 데이터들이 가지는 특성들의 범주를 일치시키기 위해 정규화를 진행하여 24개의 독립 변수로 재설정 하였다. 순환 신경망의 layer는 총 3개로 구성하였으며, 활성화 함수는 softmax를 사용하였고 최적화 함수는 adam을 사용하였다. 예측 모델의 파라미터는 학습 시 다양한 영향을 주며, 이로 인해 과대적합 및 과소적합의 문제를 야기한다. 이러한 문제를 방지하기 위해 최적의 파라미터를 찾는 과정이 필요하며, 랜덤 탐색 방법을 사용하여 최종 파라미터를 표 1과 같이 설정하였다.

표 1. 모델의 최적 파라미터

Parameter	Value
layer units	20, 20, 20
batch size	80
L2 penalty	0.07
dropout rate	0.3
time steps	2

### III. 실험 및 평가

본 논문에서 설계한 예측 모델의 성능 평가를 위해 실제 미세먼지와 예측치의 AQI 지수 비교를 수행하였다. 모델 학습을 위해 전체 데이터 중 75%를 학습 데이터로 구성하였으며, 이 중 20%의 데이터는 학습을 검증하기 위한 데이터로 사용하였다. 전체 데이터 중 남은 25%의 데이터는 모델 평가를 위해 사용하였다. 학습 횟수의 경우 학습 진행 중 결과에 대한 개선의 여지가 없을 때 학습을 종료시키는 EarlyStopping()을 이용하였다. 그림 1은 99회의 학습을 통한 모델의 loss 및 accuracy 그래프이다. 학습 모델의 validation loss는 0.62이며, validation accuracy는 0.75를 나타내었다.

구축된 모델을 통해 25%의 데이터를 이용하여 미세먼지 농도의 AQI 지수 예측을 진행하였으며, 예측 시 loss와 accuracy의 평가 수치는 0.65, 0.77의 값을 나타내었다. 표 2는 최종 예측 결과와 실제 미세먼지 AQI 지수를 비교한 결과이다. 실험 결과 각 AQI 지수 중 보통 수치의 예측 정확도가

87%로 가장 높았으며, 매우 나쁨에 해당하는 수치의 예측은 전혀 이루어지지 않는 것을 확인하였다. 매우 나쁨과 같은 결과는 나쁨 수치에 해당하는 예측 결과에서도 보여주고 있다. 이는 75%의 학습 데이터 중 나쁨과 매우 나쁨에 해당하는 데이터의 비율이 약 12%였으며, 학습을 위한 데이터양의 부족하여 생긴 문제라고 추측할 수 있다.

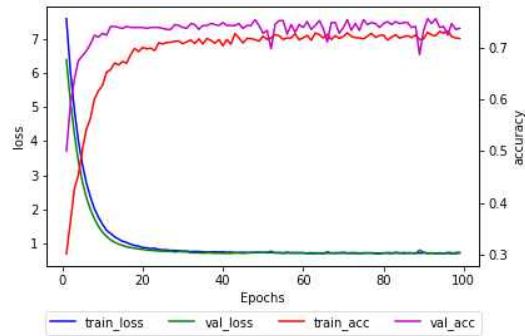


그림 1. 예측 모델 학습 결과

표 2. 예측 결과 비교

AQI 지수	실제 빈도	예측 결과		
		predict	true	accuracy
좋음	212	220	138	0.65
보통	619	655	536	0.87
나쁨	49	7	3	0.06
매우 나쁨	2	0	0	0
전체	882	882	677	0.77

### IV. 결론

본 논문에서는 인공 신경망 알고리즘 중 순환 신경망을 이용하여 미세먼지 AQI 지수 예측을 위한 모델을 설계하였다. 10년간 천안시의 일평균 기상데이터, 대기오염 물질 데이터를 활용하였으며, 랜덤 탐색 방법을 통해 최적의 파라미터를 도출하여 예측 모델에 적용하였다. 예측 모델을 통한 AQI 지수 실험 결과 나쁨과 매우 나쁨 수치에 대한 예측이 6%, 0%를 보였으며, 이는 학습 데이터 중 해당하는 수치의 양이 부족하여 생긴 문제로 판단된다. 또한 시계열 데이터의 예측을 위한 순환 신경망 알고리즘을 사용하였으나, 미세먼지의 경우 시계열 특성보다 외부 환경 및 다양한 변수에 대해 영향을 받는 불규칙적인 특성이 강한 데이터로서 순환 신경망을 이용하여 정확한 예측의 한계를 보이는 것으로 판단된다. 따라서 일평균 데이터보다 시간별 데이터를 활용함으로써 일정 수준 이상의 데이터양 확보와 외부 변수 추가 도입을 통해

예측 성능을 향상시킬 수 있을 것이며, 향후 다양한 알고리즘을 통한 예측 모델 성능 평가를 통해 미세먼지 AQI 지수 예측을 위한 최적의 알고리즘 선정을 진행할 것이다.

### References

- [1] G. W. Evans, "Air Pollution and Human Behavior," *Journal of Social Issues*, Vol. 37, No. 1, pp. 95-125, 1981.
- [2] M. S. Seo, "The Impact of Particulate Matter on Economic Activity," *The Korean Women Economists Association*, Vol. 12, No. 1, pp. 75-100, Jun. 2015.
- [3] M. Beak, and H. S. An "Empirical Leisure Environment Satisfaction Evaluation of Public Institution Employees in Innocity," *International JOURNAL OF CONTENTS*, Vol. 19, No. 2, pp. 368-378, Feb. 2019.
- [4] H. J. Yang, S. H. Kim, A. S. Jang, S. H. Kim, W. J. Song, T. B. Kim, Y. M. Y, Y. Yoo, J. H. Yu, J. S. Yoon, H. M. Jee, D. I. Suh, and C. W. Kim, "Guideline for the prevention and management of particulate matter/yellow dust-induced adverse health effects on the patients with bronchial asthma," *Journal of the Korean Medical Association*, Vol. 58, No. 11, pp. 1034-1043, Jul. 2015.
- [4] Keras Documentation [Internet]. available : <https://keras.io/layers/recurrent/>