

대기 환경 수집 IoT 장치와 가상 관측소 미세먼지 예측

최현진^{†○}, 박찬영[†], 안민지^{††}, 김효훈^{†††}, 윤재석^{†,††}

순천향대학교 사물인터넷학과 [†], 순천향대학교 미래융합기술학과^{††}, 더블유아이에스네트웍스^{†††}

{zn9852, 20191546, minji01}@sch.ac.kr, ceo@wisnet.co.kr, yun@sch.ac.kr

Particulate Matter Estimation at Virtual Stations Using Air Quality Collection IoT Devices

Hyunjin Choi^{†○}, Chanyeong Park[†], Minji An^{††}, Hyohoon Kim^{†††}, Jaeseok Yun^{†,††}

Dept. of Internet of Things[†], Dept. of Future Convergence Technology^{††}, Soonchunhyang University

WISNET Inc.^{†††}

요 약

초미세먼지는 2.5 μm 미만의 미세먼지로 세계보건기구 (WHO)에 의해 1급 발암물질로 분류되었다. 이는 인체에 깊게 침투하여 호흡기 손상 뿐만 아니라 뇌질환 등 각종 질병을 야기하기 때문에 미세먼지의 위험성을 인지하고 대비하는 것이 필수적이다. 이러한 문제의 심각성이 부각되면서 높은 정확도를 목표로 하는 미세먼지 예측 연구가 점점 중요해지고 있다. 하지만 전국에 약 600개의 미세먼지 측정소가 설치되어 있음에도 불구하고, 고가의 설치 비용 및 지리적 한계로 인해 측정 사각지대가 존재한다. 따라서 본 논문에서는 주변 측정소 데이터를 이용해 측정소가 없는 곳의 미세먼지를 예측하는 시나리오를 활용하여, 가상 관측소 (측정소가 없는 곳)의 미세먼지 예측 모델을 개발하였다. 실험을 위해 저가의 센서들로 구성된 대기 환경 수집 장치를 제작하여, 캠퍼스 내에 설치해 수집된 데이터를 활용하였다. RNN, LSTM, GRU 모델을 사용하여 24시간의 가상 관측소 미세먼지 값을 예측 및 비교 분석한 결과, LSTM 모델의 RMSE ($\mu\text{g}/\text{m}^3$)가 7.01로 가장 높은 성능을 보였다.

1. 서 론¹

초미세먼지 입자는 2.5 μm 미만의 크기로 10 μm 인 미세먼지보다도 4 배 더 작아 호흡 중에 인체의 깊은 곳까지 침투하기 쉽다[1]. 이로 인해 뇌졸중, 뇌경색이나 심장질환까지 일으키는 것으로 알려져 있다. 이러한 초미세먼지의 위험성을 인지하고 대비하기 위해서는 정확한 농도를 사전에 파악하여 인체의 악영향을 최소화하기 위한 연구가 필요하다. 에어코리아는 전국에 162 개 시, 군에 설치된 642 개의 대기환경기준물질 측정 자료를 실시간으로 제공하고 있다. 하지만 측정 장치와의 간격이 커짐에 따라 측정 사각지대가 발생하기 때문에 정확한 미세먼지 농도를 파악하기 어렵다. 또한 대기 환경 측정 장비는 고가의 측정 장비이기 때문에 측정소를 무작정 늘릴 수도 없는 상황이다.

이에 따라 미세먼지 예측에 대한 다양한 연구가 활발하게 진행되고 있다. An 등은 IoT 기반 PM 수집 시스템을 제안하고, 에어 코리아 자료를 이용하여 가상 관측소의 PM2.5 농도를 예측해 봄으로써, 주변 측정소의 대기 오염 인자 데이터를 이용하여 가상 관측소의 PM2.5 농도를 예측할 수 있다는 가능성을 보여주었다 [2]. Inapakurth 등은 대기 중 입자 물질 및 오염 물질인 15 개의 특성을 사용하여 딥러닝

모델을 생성하였고 다중 목표 최적화를 통해 정확도와 복잡성 균형을 맞추는 신경망을 찾는 알고리즘을 제안하였다 [3]. 또한 최근 장기 예측에 대한 관심도 높아지고 있다. Jeong 등은 기후 변수들의 연관성을 활용하여 미세먼지 농도를 수개월 전에 장기 예측이 가능한 계절 예측 기법을 제안하였다 [4].

기존 연구들은 주로 특성들 간의 연관성을 활용한 최적화에 중점을 둔 것과 달리 본 논문은 측정 장비가 없는 지역에서도 기존 측정 장비를 활용해 미세먼지 예측을 위한 데이터 수집과 분석을 가능하게 하는 접근 방식을 제안한다. 선행 연구에서 제안한 IoT 기반 PM 수집 장치를 도시 대기 측정망 측정소 설치 기준과 유사하게 캠퍼스 내에 배치하고, 4 개의 장치 중 한 곳을 센서가 없는 가상 관측소라고 가정하여 가상관측소 PM2.5 값을 예측하는 모델을 비교 분석하였다.

2. 시스템 구현

2.1 LoRaWAN 연동 대기 환경 수집 장치

대기 환경 수집 장치에 사용된 보드는 Adafruit Feather M0 with LoRa 보드이며, 수집 장치의 구성은 그림 1과 같다. 미세먼지 센서는 실시간 측정이 가능한 광산란 방식의 PMS7003 센서를 사용하였다. 또한 온습도, 일산화탄소, 이산화질소, 오존 센서를 추가로 설치하여, 미세먼지와 다른 대기 환경 인자와의 상관 관계 분석을 통한 모델 학습이 가능하도록 설계하였다.

– 본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구 결과로 수행되었음(2021-0-01399)

– 이 연구는 2024년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임 (RS-2022-00154678)

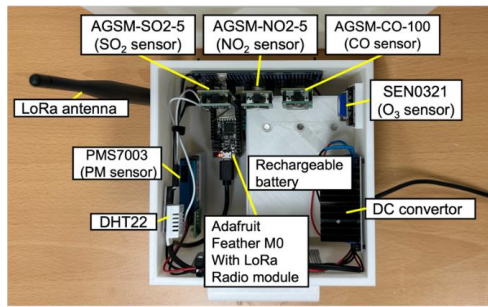


그림 1. 대기 환경 수집 장치 구성

2.2 시스템 구성

수집 장치의 각 센서는 여러 통신 방식 (UART, SPI, I2C)을 활용하여 측정되고, 장거리 및 저전력 통신인 LoRa를 사용하여 5분마다 데이터 송수신이 진행된다. 여러 장치에서 동시에 수집되는 데이터의 패킷 누락을 최소화 하기 위해, 8채널 LoRaWAN 게이트웨이인 RAK2245 Pi HAT을 사용하였다. 수집된 데이터는 LoRaWAN 네트워크 서버를 거쳐 oneM2M 기반 오픈 소스 플랫폼인 모비우스에 업로드 하였다.

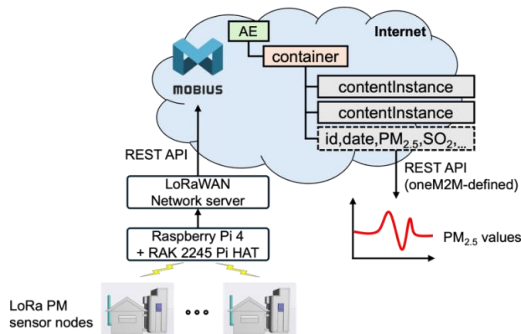


그림 2. 시스템 아키텍처

2.3 장치 배치 및 실험 설계

장치 배치는 그림 3과 같으며, 게이트웨이는 그림 3의 빨간 위치에 설치하였다. 에어코리아의 도시 대기 측정망 측정소 설치 기준과 유사하도록 건물 옥상에 설치하여 24년 2월 ~ 3월까지 한 달동안 데이터 수집을 진행하였다. 가상 관측소는 자동차의 통행이 없고, 유동인구가 많은 곳인 그림3의 2의 위치에 설치된 장치로 선정하여 실험을 진행하였다.



그림 3. 장치 배치 및 게이트웨이 위치 (출처: 네이버)

3. 예측 모델 설계

3.1 데이터 수집 및 전처리

캠퍼스 내에 배치한 대기 환경 수집 장치를 통해 2024년 2월 18일 20시부터 3월 20일 11시까지 수집된 데이터를 활용하여 가상 관측소 PM2.5 예측 모델 학습 및 성능 분석을 진행하였다. 패킷 누락과 배터리 방전 등으로 인한 결측치는

이전 값과 이후 값의 직선 거리에 따라 선형적으로 계산하여 추정하는 방법인 선형 보간법을 사용하였다. 그림 4는 선형 보간법을 통해 결측치를 처리한 후의 미세먼지 값을 선 그래프로 나타낸 모습이다. 모델 학습을 위한 스케일링은 MinMaxScaler를 사용하였고, 전체 데이터의 80%는 훈련셋으로, 20%는 테스트셋으로 구성되고, 훈련셋 중 20%는 검증셋으로 구성하였다.

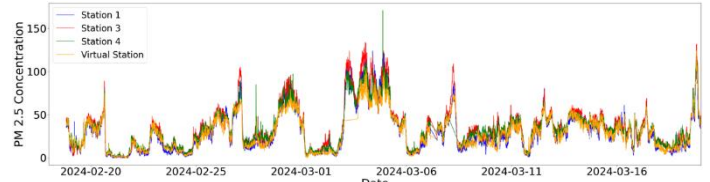


그림 4. 수집된 초미세먼지 데이터

3.2 실험 방법 및 결과 분석

3개의 장치 (그림3-1,3,4)의 데이터로 가상 관측소 (그림3-2)의 PM2.5 값을 예측하기 위해 주로 시계열 모델 예측에 쓰이는 대표적인 딥러닝 모델인 RNN, LSTM, GRU를 사용하였다. RNN은 순차 데이터를 처리하기 위해 순환하는 구조를 가지고 있으며, LSTM과 GRU는 RNN의 장기 의존성 문제를 해결한 모델이다[5]. RNN, LSTM, GRU 모델에서 하이퍼파라미터 튜닝을 통해 찾아낸 최적의 파라미터 값은 아래와 같다. 각 모델은 그림 5-a, 5-b, 5-c와 같이 학습 곡선을 통해 학습 여부를 확인하였다.

Parameter	RNN	LSTM	GRU
Input shape	(47,288,18)	(47,288,18)	(47,288, 18)
Units	32	32	16
Learning rate	0.01	0.01	0.01

표 1. 각 모델 하이퍼파라미터 값

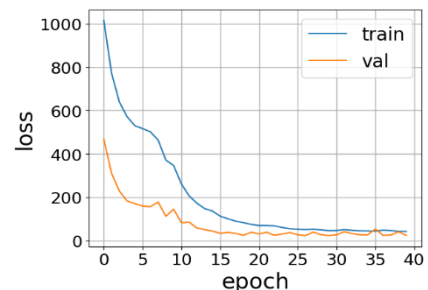


그림 5 - a. RNN 학습 곡선

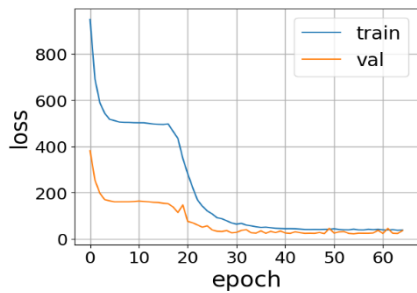


그림 5 - b. LSTM 학습 곡선

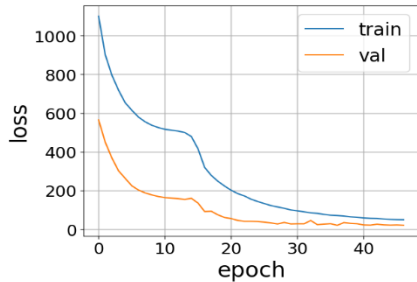


그림 5 - c. GRU 학습 곡선

그림 6은 장치에서 수집한 데이터를 이용하여 3월 14일 9시 이후 24시간동안 가상관측소의 PM2.5 값을 예측한 그래프이다. 세 모델은 모두 비슷한 추세를 보였으며, 실제 관측 값의 추세와 비슷함을 볼 수 있다. 하지만 14일 9시부터 15시 사이 특정 시간대에서는 세 모델의 예측 값이 실제 값과 큰 차이가 나는 것을 볼 수 있었다. 각 모델을 비교해보면, 표 2의 RMSE 값에서 LSTM 모델이 7.01로 가장 낮게 나타나 세 모델 중 가장 높은 성능을 보였다. LSTM 모델은 값의 변화가 큰 시점에도 비교적 정확하게 예측하였으며, RNN 모델은 대체적으로 실제 값과 비슷한 그래프가 나타났으나 변화가 큰 시점에서 제대로 예측이 되지 않은 것으로 보였다.

Case	RNN	LSTM	GRU
RMSE ($\mu\text{g}/\text{m}^3$)	7.62	7.01	7.87

표 2. 가상 관측소 PM2.5 예측 RMSE 결과 비교

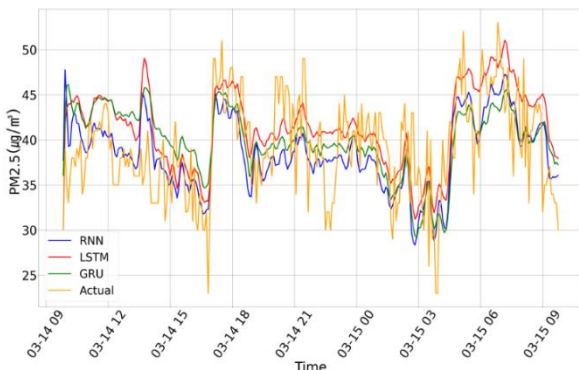


그림 6. 가상 관측소 PM2.5 예측 그래프

그림 7은 위의 결과에서 가장 성능이 우수했던 LSTM 모델을 이용하여 예측한 값과 관측소 1, 관측소 3, 관측소 4의 미세먼지 값을 그래프로 나타내었다. 관측소 1은 PM2.5 대기환경 기준으로 좋음 ($0 \sim 15 \mu\text{g}/\text{m}^3$) 범주에 해당하는 일자가 많은 것으로 나타났으며, 관측소 3과 4는 보통

($16 \sim 35 \mu\text{g}/\text{m}^3$)에서 나쁨 ($36 \sim 75 \mu\text{g}/\text{m}^3$) 기준의 농도로 주로 이루어져 있음을 볼 수 있었다. 가상 관측소의 예측 값은 관측소 1의 값과 유사한 수준을 보이며, 관측소 3과 4의 값과는 약간의 차이가 있음을 확인할 수 있었다.

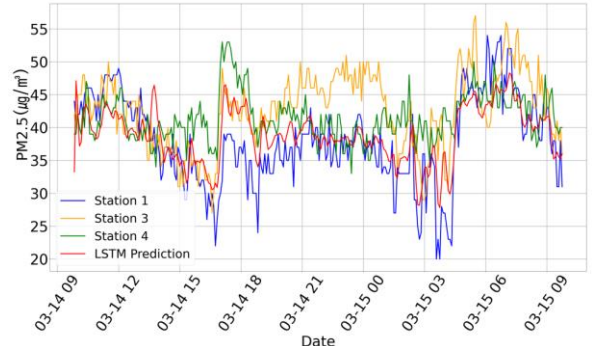


그림 7. 관측소 1, 3, 4와 PM2.5 예측 비교 그래프

4. 결론

특성들 간의 연관성을 활용한 최적화에 중점을 둔 기존 연구와 달리, 본 연구에서는 측정소가 없는 지역에서도 주변 장치를 이용해 데이터 수집 및 분석함으로써 PM2.5 예측이 가능함을 보여주었다. 선행 연구에서 제안한 IoT 기반 대기 환경 데이터 수집 장치를 캠퍼스 내에 배치하고, 수집한 데이터를 활용하여 딥러닝 모델을 통해 PM 2.5 농도를 예측하였다. 그 결과, 세 모델 예측 값의 추세가 실제 값과 비슷한 것을 보아 3개의 관측소의 데이터를 통해 가상 관측소의 PM2.5 값 예측이 가능함을 확인하였다. 또한 세 모델 중에서 LSTM 모델의 RMSE 값이 7.01로 가장 우수한 성능을 보였으며, 기존 측정소의 대기 환경 데이터를 활용하여 측정소가 없는 지역의 미세먼지 농도를 보다 정확하게 예측할 수 있음을 입증하였다. 이러한 결과를 토대로, 향후 가상관측소를 다른 위치로 변경해가며, 주변 장치를 이용한 가상 관측소의 PM2.5를 예측하는 연구를 진행할 예정이다.

5. 참고 문헌

- [1] 이득우, "초미세먼지 농도 예측을 위한 랜덤 포레스트 기반 예측 모델." 국내석사학위논문 숭실대학교 대학원, 2020
- [2] 안민지, 박찬영, 노현정, 윤재석, "IoT 기반 PM 수집 시스템과 예측 방법", 한국정보과학회 2023 한국컴퓨터종합학술대회 논문집, 170-172, 2023
- [3] Inapakurthi, Ravi Kiran et al. "Deep learning based dynamic behavior modelling and prediction of particulate matter in air." Chemical Engineering Journal 426, 2021
- [4] Jeong, Jee-Hoon, et al. "A novel statistical-dynamical method for a seasonal forecast of particulate matter in South Korea." Science of The Total Environment 848, 2022
- [5] Lindemann, Benjamin, et al. "A survey on long short-term memory networks for time series prediction." Procedia Cirp 99, 2021