

## 미세먼지 종류의 구분이 가능한 분류모델

전민혜<sup>1</sup>, 안석호<sup>2</sup>, 서영덕<sup>3</sup><sup>1</sup>인하대학교 컴퓨터공학과 학부생<sup>2</sup>인하대학교 컴퓨터공학과 석사과정<sup>3</sup>인하대학교 컴퓨터공학과 교수

pushpull7381@gmail.com, sokho0514@gmail.com, mysid88@inha.ac.kr

## A Classification Model for Distinguishing the Types of Fine-dust

Minhye Jeon<sup>1</sup>, Seokho Ahn<sup>2</sup>, Young-Duk Seo<sup>3</sup><sup>123</sup>Dept. of Computer Science and Engineering, Inha University

## 요약

대기오염에 대한 사회적 관심이 증가함에 따라 미세먼지 관련 IoT 가전제품의 수요가 증가하였다. 이에 따라 IoT 제품에서 수집되는 실내 데이터의 정확성을 높이기 위해 미세먼지 센서 외의 추가적인 센서를 사용하여 관측치를 예측하는 연구가 진행되고 있다. 하지만 기존 연구에서는 센서의 물리적 한계를 인공지능을 통하여 극복하려는 시도는 존재하지 않는다. 본 논문에서는 추가적인 센서를 사용하지 않고 단일 센서만을 사용하여 미세먼지 종류의 구분 가능성을 판단하고자 하며 이를 정확도 기반으로 비교 실험을 진행하여 가장 좋은 성능을 나타낸 딥러닝 기반 분류모델을 선정한다.

## 1. 서론

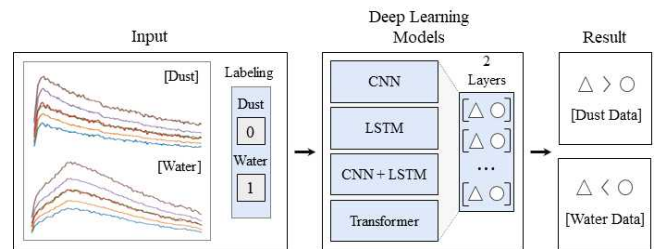
오늘날 대기오염에 대한 사람들의 인식이 높아져 미세먼지 센서는 일상생활 속에서 다양한 용도로 쓰이고 있다. 이에 미세먼지 관련 IoT 가전제품의 수요 또한 증가하고 있다. 따라서 미세먼지 센서 기반 가전제품의 성능향상을 위해 효율성과 출력값에 대한 정확도를 높이기 위한 노력이 필요하다.

그러나 시장에 주로 보급되는 저가형 센서는 단순 입자의 크기만으로 미세먼지의 종류를 파악하기 때문에 만일 습도가 높은 경우, 저가형 센서는 수분 데이터의 입자를 미세먼지로 인식하여 미세먼지 수치에 대한 정확도가 낮아지는 현상이 나타난다 [1]. 이러한 문제를 해결하기 위해 온·습도 센서와 미세먼지 센서로부터 수집한 데이터를 학습하여 미세먼지 농도를 예측하는 연구들이 수행되고 있다 [2]. 그러나 해당 데이터는 온·습도 센서의 사용이 필수적이며, 미세먼지 센서 자체만 사용하는 연구는 존재하지 않는다. 단일 센서만을 사용하여 얻은 미세먼지 데이터의 종류를 구분하여 예측의 정확도를 높이는 방안이 필요하며, 이는 센서의 물리적인 한계를 극복하는 방안이라는 점에서 의의가 있다.

본 논문에서는 단일 미세먼지 센서를 통해 미세먼지 데이터를 수집하고, 딥러닝 기반 분류모델을 통해 미세먼지 종류의 구분이 가능함을 보이고자 한

다. 만일 단일 센서만으로 미세먼지 종류 구분이 가능할 경우 정확성 향상과 동시에 상대적으로 저렴한 IoT 가전 기기를 제공할 수 있을 것이다.

## 2. 미세먼지 구분이 가능한 딥러닝 기반 분류모델



(그림 1) 미세먼지 종류의 구분이 가능한 딥러닝 기반 분류모델

본 논문에서는 미세먼지와 수분 데이터를 구분하는 것에 초점을 맞춘다. 추가적인 온·습도 센서를 사용하지 않고 미세먼지 센서만 사용하여 수집된 시계열 데이터를 통해 미세먼지 종류를 분류할 수 있는 최적의 딥러닝 기반 분류모델을 찾는 것을 목표로 하며, 제안하는 분류모델은 그림 1과 같다.

실험을 위해 단일 미세먼지 센서만을 사용하여 수분 데이터와 미세먼지 데이터를 수집한 뒤, 데이터의 전처리 및 라벨링을 진행하는 과정을 거친다. 이후 분류 문제에서 많이 사용되는 딥러닝 알고리즘인 CNN, LSTM, CNN과 LSTM 결합 모델

(CNN+LSTM), Transformer 총 4가지의 분류모델을 적용하여 분류의 정확성을 측정하였다 [3]. 제안 모델은 이진 분류를 사용하기 때문에 두 개의 출력값이 도출된다. 해당 값은 0과 1 사이의 값으로 이루어져 있으며, 0에 가까울수록 미세먼지로, 1에 가까울수록 수분 데이터로 분류한다. 이후 정답 데이터와 비교하여 성능을 측정한다.

**2-1. 데이터셋과 전처리**

본 논문에서는 총 4개의 레이저 방식 미세먼지 센서(PMS7003)를 통해 일반 가정환경에서 발생할 수 있는 다양한 미세먼지 데이터와 가습기로부터 수집한 수분 데이터로 학습을 진행하였다. 레이저 센서는 미세먼지 입자의 크기 별 측정이 가능하므로, 충분한 데이터가 수집될 경우 분류 모델을 통해 미세먼지 종류별 구분이 가능하다고 판단하였다.

수집된 데이터의 결측치와 이상치는 각각 선형 보간법과 IQR (Interquartile range)을 통해 처리하였다. 또한, 센서 측정 오류나 통신 문제 등으로 달라진 수집 간격을 10초 간격으로 동기화하였으며, 데이터의 노이즈는 HP 필터와 SMA 기법을 통해 제거하였다.

**2-2. 실험 구성**

본 논문에서는 수집된 데이터로부터 미세먼지 데이터 수치의 변화율과 수분 데이터 수치 변화율의 차이에서 발견되는 시계열 데이터의 학습을 진행하였으며, 이를 총 4가지의 딥러닝 기반 분류모델을 구축하여 비교 실험을 진행하였다. 입력 데이터는 PM1.0, PM2.5, PM10을 CF, AT로 측정하여 6개의 특징을 병렬로 설정하였다. 여기서 CF는 표준대기압을 상정했을 때 측정된 미세먼지 수치를, AT는 현재 대기 환경에서의 미세먼지 수치를 의미한다. 이렇게 병렬로 설정된 데이터의 윈도우 크기는 10으로 지정하여 각 모델이 과거 100초 간격의 데이터를 학습할 수 있도록 하였다. 이를 9:1로 나누어 학습용 데이터와 검증용 데이터를 구성하였고 학습률(Lr)은 0.001과 0.0001 두 가지로 나누어 실험을 진행하였다.

**2-3. 실험 결과**

<표 1> 분류모델 비교 평가

Lr	0.001		0.0001	
	Acc	F1	Acc	F1
<b>CNN</b>	0.895	0.932	0.851	0.905
<b>LSTM</b>	0.879	0.924	0.902	0.936
<b>CNN+LSTM</b>	0.953	0.967	<b>0.906</b>	<b>0.939</b>
<b>Transformer</b>	<b>0.963</b>	<b>0.976</b>	0.738	0.845

본 논문에서는 Accuracy (Acc)와 F1 score를 평가 지표로 사용하여 평가를 진행하였으며 실험 결과는 표 1과 같다. 먼저 학습률을 0.001로 설정한 경우, Transformer를 사용한 분류모델이 가장 좋은 성능을 보임을 확인하였다. 반대로 학습률을 0.0001로 설정하였을 경우, Transformer를 사용한 모델은 낮은 학습률 설정으로 인하여 모델의 가중치 업데이트가 느려져 낮은 정확도를 보였다. LSTM, CNN+LSTM을 사용한 각 모델에서의 정확도는 이전 실험보다 비교적 안정적으로 증가하는 현상을 관측하였고 결과적으로 CNN+LSTM을 사용한 분류모델이 가장 좋은 성능을 보임을 확인하였다.

**3. 결론**

본 논문에서는 미세먼지와 수분 데이터를 구분하기 위한 분류모델의 비교 실험을 진행했다. 또한, 비교 모델의 성능을 확인하기 위해 다양한 비교분석을 수행하였다. 실험 결과, 딥러닝 알고리즘을 통하여 추가적인 센서가 없어도 높은 정확도로 미세먼지와 수분 데이터를 구분할 수 있음을 증명하였다. 향후 연구로는, 본 논문에서 성능이 가장 좋은 분류모델을 참고하여 단일 센서를 사용한 데이터 기반으로 미세먼지 종류 구분에 특화된 새로운 모델을 제안하고자 한다. 또한, 낮은 하드웨어 환경에 제안하는 모델을 탑재하여 실제 IoT 가전제품에서의 성능을 측정해보고자 한다.

**사 사**

본 연구는 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2022R1C1C1012408, 우수신진연구)과 정보통신기획평가원의 지원(No.2022-0-00448, 사람중심인공지능핵심원천기술개발, No.RS-2022-00155915, 인공지능융합현신인재양성(인하대학교))을 받아 수행한 연구임.

**참고문헌**

[1] 이창호, “대기 습도가 광 산란 미세먼지 측정에 미치는 영향.”, J. Korean Soc. Hazard Mitig 2020; 20(1): 391-399.

[2] Chuluunsaikhan, T.; Heak, M.; Nasridinov, A.; Choi, S. Comparative Analysis of Predictive Models for Fine Particulate Matter in Daejeon, South Korea. Atmosphere 2021, 12, 1295. <https://doi.org/10.3390/atmos12101295>

[3] Lu, Wenjie, et al. "A CNN-LSTM-based model to forecast stock prices." Complexity 2020 (2020): 1-10.