

학교 미세먼지 예측 모델 개발

정종덕¹, 한규한¹, 김성석², 김재현³

¹서경대학교 소프트웨어학과

²서경대학교 소프트웨어학과, 교신저자

³서경대학교 컴퓨터공학과

jjduck0307@naver.com, {gksrbgks713, sskim03, statsr}@skuniv.ac.kr

Development of Prediction Model for School Fine Dust

Jongduck Jeong¹, Gyoohan Han¹, Sunguk Kim², Jaehyun Kim³

¹Dept. of Software Engineering, Seokyeong University

²Dept. of Software Engineering, Seokyeong University, Corresponding author

³Dept. of Computer Engineering, Seokyeong University

요 약

세계보건기구에 의해 미세먼지가 1급 발암물질로 지정됨에 따라, 건강에 미치는 영향에 대한 관심이 커지고 있다. 성장기 학생들은 학교에서 오랫동안 머무르게 되며, 따라서 교내 미세먼지는 학생들의 건강에 미치는 영향이 매우 클 수 있다. 본 연구에서는 1년동안 대상 학교에서 수집한 미세먼지 관련 데이터 및 공개된 Air Korea, 기상청 데이터를 활용하여, LSTM 기반 미세먼지 예측 모델을 개발하였다. 수집한 데이터에서 일부 오류나 결측치가 있었지만, 이를 활용한 미세먼지 예측 정확도는 충분히 가치가 있다고 판단된다.

발하고자 한다.

1. 서론

미세먼지는 오래전부터 환경 분야에서 큰 관심을 받아온 주제이다. 우리나라는 국내적인 원인뿐만 아니라 해외에서 발생한 미세먼지에도 큰 영향을 받고 있다. 미세먼지는 지름이 $10\mu\text{m}$ 인 물질로서(PM_{10}), 세계보건기구에서 1급 발암물질로 지정되었다[1]. 최근에는 그보다 더 크기가 작은, 지름 $2.5\mu\text{m}$ 이하인 미세먼지($\text{PM}_{2.5}$)가 인체에 미치는 영향이 더 큰 것으로 밝혀져 많은 연구가 진행되고 있다.

인체가 일정 기간동안 미세먼지에 노출되면, 단순히 폐렴이나 만성폐쇄성폐질환에서부터 폐암까지 발생시키는 것으로 보고되고 있다. 어린이들은 건강 취약계층으로 분류되고 있으며, 따라서 미세먼지에 오랫동안 노출될 경우 성장기에 여러 악영향을 미치는 것으로 보고되고 있다. 과거 미세먼지 관련 연구들은 주로 넓은 지역에서 미세먼지가 미치는 영향 분석에 초점을 맞추었다[2, 3]. 이에 반해 본 연구에서는 어린 학생들이 오랫동안 머무르는 초등학교나 중학교 실내 미세먼지를 예측할 수 있는 모델을 개

2. 미세먼지 예측 모델

본 연구는 참여중인 사업단에서 대상 학교(조현초와 중대부중)에서 일정 기간(1년)동안 측정하고 수집한 교내(내기)와 운동장(외기) 미세먼지 데이터와, 한국환경공단에서 제공하는 Air Korea 데이터, 그리고 기상청 데이터를 활용하였다[4, 5]. Air Korea 데이터는 학교와 가장 가까운 위치의 측정소에서 수집한 데이터를 활용하였고, 기상청 데이터는 기상자료개방포털에서 학교가 속한 동 단위 데이터(바람의 세기와 방향)를 활용하였다.

본 연구에서는 여러 데이터 중에서 미세먼지 PM_{10} 와 $\text{PM}_{2.5}$ 를 예측하는 것에 초점을 맞추었다. 우선 학교에서는 매 2분마다 데이터를 측정하였으며, Air Korea와 기상청은 수집한 데이터를 전처리하여 매 1시간마다 측정된 값을 제공하고 있다. 이처럼 측정 주기가 다른 데이터를 함께 고려하기 위하여, 수집한 데이터에서 1시간 단위로 측정된 값을 활용하였다((그림 1) 참고).

Air Korea 데이터는 대개 공공기관 옥상에 설치된 측정장비에서 수집하며, 따라서 지상 2-3층 높이의 미세먼지 데이터가 된다. 이에 반해 학교 운동장

* 이 논문은 정부(과학기술정보통신부, 교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2019M3E7A1113102).

에서 수집한 데이터는 지상 1m 높이에 센서가 수집한 데이터가 된다.

	ak_PM10	ak_PM25	speed	direction	31_PM10	31_PM25
2020-07-01 08:00:00	14.0	4.0	0.7	54	12.066667	2.333333
2020-07-01 09:00:00	14.0	4.0	0.9	208	11.700000	2.200000
2020-07-01 10:00:00	15.0	7.0	0.3	230	11.900000	2.333333
2020-07-01 11:00:00	16.0	8.0	0.5	219	12.566667	2.733333
2020-07-01 12:00:00	15.0	8.0	0.1	194	16.517241	3.379310
...
2020-12-31 19:00:00	34.0	17.0	0.5	211	11.533333	7.266667
2020-12-31 20:00:00	32.0	15.0	0.4	215	15.033333	9.066667
2020-12-31 21:00:00	34.0	17.0	0.5	216	13.896552	8.931034
2020-12-31 22:00:00	31.0	18.0	0.6	229	12.333333	6.933333
2020-12-31 23:00:00	33.0	18.0	0.2	228	12.866667	9.266667

(그림 1) 활용한 데이터 샘플 ('ak_'는 Air Korea 데이터)

미세먼지 예측을 위해 기본적으로 Air Korea 데이터를 이용하여 학교 운동장의 미세먼지를 예측하는 LSTM 모델을 개발하고자 한다. 학교 운동장에서 수집한 데이터는 기계학습에서 활용하였다. 학습을 위해 전체 80% 데이터를 이용하였고, 20%는 검증용으로 사용하였다. 현재 학교에서 수집한 데이터에 일부 구간에서 결측치가 존재하여 이를 전처리하여 제거하였다.

```

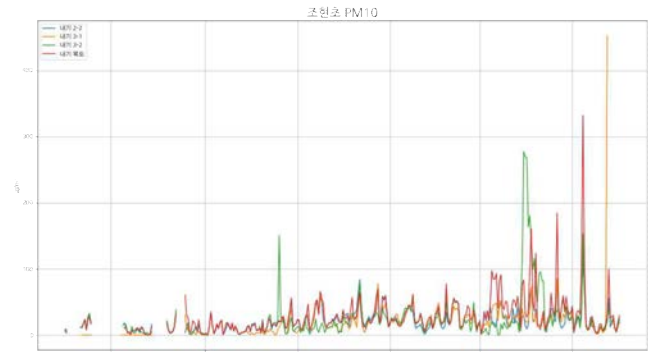
model.add(LSTM(16,
               dropout=0.2,
               recurrent_dropout=0.2,
               input_shape=(train_feature.shape[1], train_feature.shape[2]),
               activation='relu',
               return_sequences=False))
    
```

(그림 2) 피이션으로 구현한 LSTM 기반 모델

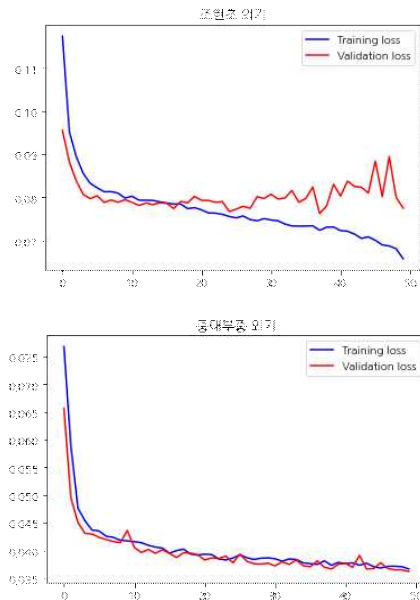
먼저 (그림 3)은 조현초에서 수집한 학교 데이터의 운동장과 실내 미세먼지와 관계를 표현하고 있다. 실내의 PM₁₀ 미세먼지 값은 유사한 부분도 있지만, 그렇지 않은 구간도 존재한다. 이는 학생들이 실내에서 활동을 하거나 공조기를 가동시킬 경우, 실내 미세먼지 값이 운동장(외기)과 차이가 발생할 수 있기 때문이다. 그림에는 없지만, 중대부중도 유사한 경향을 보이지만 일정 구간동안 이상값이 존재하여 이를 처리하여 사용해야 했다.

(그림 4)은 앞서 이용한 LSTM 모델에 의해 Air Korea 데이터를 이용하여 학교 운동장의 미세먼지 (PM₁₀)를 예측한 결과를 보여주고 있다. 그림에서 조현초는 학습 결과 예측값의 오차율이 0.07% 수준이었으며, 중대부중은 학습 결과에 의한 예측 정확도가 상대적으로 더 좋게 나왔다(0.035%). 이처럼 외기 데이터에 대한 예측이 가능하다면, 교실 내에

서 학생들의 활동이나 공조기 가동 여부를 함께 고려할 때 미세먼지 양을 좀 더 정확하게 예측이 될 것으로 판단된다.



(그림 3) 실내(내기) 데이터와 운동장(외기) 데이터 관계



(그림 4) LSTM 모델에 의한 예측 결과

3. 결론

본 연구에서는 학교 내에서 미세먼지 예측을 위한 모델을 개발하고 검증한 결과를 담고 있다. 공개된 Air Korea 및 기상청 데이터와 대상 학교에서 측정된 데이터를 활용하여 LSTM 기반 모델을 개발하였다. 아직은 데이터의 검증 단계를 거쳐야 하지만, 어느 정도의 정확도를 보여줄 수 있는 것으로 판단된다.

참고문헌

[1] M. Santibanez-Andrade, Y.I. Conzalez-Ramirez, Y. Sanchez-Perez, and C.M. Garcia-Cuellar, "Deciphering the code between air

pollution and Disease: The effect of particulate matter on cancer hallmarks”, Journal of Molecular Sciences, 21(1), pp. 1-15, 2020.

[2] Hong-jin Park, "Analysis and Prediction of (Ultra) Air Pollution based on Meteorological Data and Atmospheric Environment Data”, Journal of Korea Institute of Information, Electronics, and Communication Technology, 14(4), pp. 328-337, 2021.

[3] Dong-jun Won, Sun-Kyun Kim, Yeonghun Kim, Syuwon Song, "Prediction of fine Dust in Gyeonggi-do Industrial Complex using Machine Learning Methods”, Journal of Korea Institute of Information Scientists and Engineers, 48(7), pp. 764-773, 2021.

[4] <http://210.110.39.121/padmin/index.php>

[5] <https://www.airkorea.or.kr>