

효율적인 도로 비산먼지 제거 경로 제안을 위한 LSTM 기반 미세먼지 예측

임동진*, 김태홍**, 이용**, 정한민**

*강원대학교 컴퓨터정보통신공학과

**한국과학기술정보연구원

e-mail : glow147@naver.com, {thkim,ryonglee,jhm}@kisti.re.kr

LSTM-based Particulate Matter prediction for efficient road scattering dust removal path proposal

DongJin Lim*, Taehong Kim**, Ryong Lee**, Hanmin Jung**,

*Dept. of Computer and Communications Engineering, Kangwon National University

**Korea Institute of Science and Technology Information

요 약

1 급 발암물질인 미세먼지 중 44.3%를 차지하고 있는 도로 비산먼지는 효과적인 미세먼지 농도 저감 대책의 방안 중 하나이다. 도로 비산먼지 제거는 일반적으로 특수 차량을 이용, 정해진 경로와 주기에 따라 운행된다. 이러한 운행방식은 도로의 오염 현황에 따른 효과적 경로 선정 및 운영이 어렵다. 본 논문에서는 도로 비산먼지 제거의 효율적인 경로 제안을 위해 대구지역에 분포된 KISTI 이동형 도시센싱 테스트베드에서 수집되는 고해상도의 실시간 지역별 오염 현황 데이터를 활용하여 실시간 오염도를 분석하고, LSTM(LONG SHORT-TERM MEMORY) 알고리즘을 활용하여 미래의 미세먼지 농도를 예측하였다. 기존 연구와 달리 지역별 상황을 고려한 데이터를 사용하여 선형 회귀 분석을 수행하였다. 실험 결과, 시간 속성을 고려한 LSTM 이 MLP 보다 평균 제공근 오차 값이 경우에 따라 최대 30% 더 작음을 확인했다. 본 연구를 기반으로 고해상도 사물 데이터 기반 예측 연구의 가능성을 보였으며, 미세먼지 예측 결과를 활용 유연하고 효과적인 도로 청소차량의 운행 경로를 설정에 활용될 수 있을 것으로 기대한다.

1. 서론

미세먼지의 심각성이 지속적으로 대두되고 있는 가운데 환경부에서는 미세먼지 저감을 위해 노후경유차량 조기폐차, 도로 청소차량 보급, 전기차 보급 등의 저감 대책방안을 마련하고 있다. 이 중 도로 청소차량은 도로 비산먼지를 제거하는 차량으로 도로에 쌓인 먼지가 재비산 되는 것을 막는다. 도로 재비산먼지는 2013 년 전국 미세먼지(PM10) 배출량 통계에서 44.3%로 가장 높은 비율을 차지한다[1]. 도로 재비산먼지는 Aluminum, Kalium 등과 같은 지각 물질에 기인하는 자연적 성분 외에도 배출가스, 타이어 및 브레이크 마모 등에 의해 발생하는 Cadmium, Lead, Chromium 등 유해한 인위적 성분을 포함하여 일반 먼지보다 더욱 인체에 해롭다. 도로 비산먼지는 특히 교통·산업·건설 활동이 많은 대도시에서 많이 발생한다[2]. 이처럼 도로 비산먼지 제거 사업은 도시민의 건강을 위해 중요한 부분이다. 최근 연구에 따르면 미세먼지(PM2.5 · 지름이 2.5 μm 이하 입자상태)를 1 ton 감축하는데 가장 비용이 적게 들어가는 방안은 도로에 물을 뿌려 청소하는 방법(500 만원)으로 나타났다[3]. 도로 청소 전과 후의 미세먼지 농도 변화를

조사한 연구에서는 특히 PM10 이상의 먼지 저감 효과가 뛰어난 것으로 조사됐다[4]. 하지만 현재 도로 청소차량은 서울시를 제외하면 모든 도로를 청소하기에 차량 수가 부족하다. 경기도의 경우 도로의 길이가 서울 보다 1 만 1 천 552km 더 길지만 도로 청소차량의 보유대수는 서울시 보유량의 73% 수준에 불과하다[5]. 이처럼 대부분의 지자체들이 도로 길이에 비해 차량 보유량이 부족하다. 따라서 효율적인 도로 비산먼지 제거 경로가 필수적이다. 효율적인 경로 제안을 위해서는 지역별 오염 현황을 알아야하기 때문에 고해상도의 지역 오염 데이터가 필요하다.

이를 위해서 본 논문에서는 KISTI 이동형 도시센싱 테스트베드에서 17 년 6 월 8 일부터 수집하고 있는 데이터를 활용하였다. 수집되는 데이터는 <표 1>과 같으며, 이 데이터를 활용하여 미래의 PM10 농도를 Keras 환경에서 딥 러닝을 통해 예측하고자 한다.

<표 1> KISTI IoT 센싱 데이터 정보

Section	Example value	단위	Section	Example value	단위
시간	2017-04-27 07:13:01	년-월-일 시:분:초	NO2[이산화질소]	0.049	ppm
경도	128.60176	Degree	SO2[이황산화가스]	0.052	ppm
위도	35.87141	Degree	CO[일산화탄소]	0.35	ppm
온도	27	°C	VOC	1	ppm
습도	98	%	대기압	1013	hPa
PM2.5	27	µg/m³	소음	54	dBA
PM10	98	µg/m³	진동(자이로)	49.78.72	mg(x,y,z)

2. 관련연구

2.1 환경부 도로비산먼지 측정

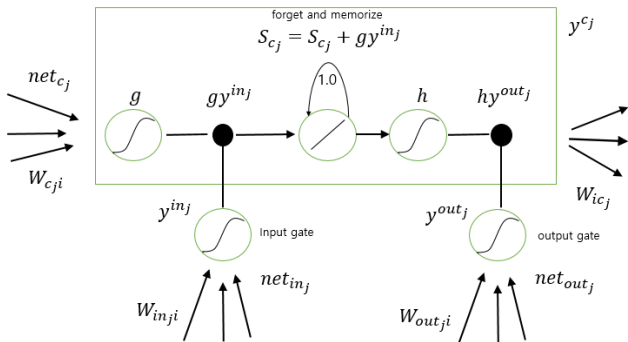
환경부에서는 효율적인 도로 비산먼지 제거를 위해 도로먼지 측정 차량을 운행 중 이다. 하지만 아직 수도권 일부만 측정 중 이며 보유 차량이 적어 모든 지역을 측정할 수 없다. 본 논문에서는 택시가 이동하면서 수집하는 센서 데이터를 이용하여 이 문제를 해결 하였다.

2.2 Machine Learning 을 이용한 미세먼지(PM10) 농도 예측

신혜영[6]과 오병두 외 2 명[7]이 여러가지 기계학습 기법들을 이용한 알고리즘 성능 측정 및 미세먼지 예측을 연구하였다. 해당 연구들에서 Multi Layer Perceptron(MLP)의 결과가 가장 좋았으며, 본 논문에서는 시간 속성을 고려한 LSTM 알고리즘을 사용하여 MLP 알고리즘과 성능을 비교하였다.

3. 학습 알고리즘

LSTM 은 RNN 의 장기기억문제를 개선한 알고리즘으로 아래(그림 1)와 같은 구조를 갖는다. [8]



(그림 1) feedback with 1 time step

c_j 는 j 번째 메모리 셀을 의미하며, 는 메모리 셀의 내부 상태를 나타낸다. 각각의 입력 값들은 다음과 같다.

$$net_{out_j}(t) = \sum_u w_{out_j,u} y^u(t-1)$$

$$net_{in_j}(t) = \sum_u w_{in_j,u} y^u(t-1)$$

$$net_{c_j}(t) = \sum_u w_{c_j,u} y^u(t-1)$$

이러한 다양한 유형의 유닛들은 현재 신경망의 상태에 대한 유용한 정보를 전달할 수 있다. 예를 들어

입력 게이트(출력 게이트)는 다른 메모리 셀로부터의 입력을 사용하여 메모리 셀에 특정 정보를 저장할지 여부를 결정할 수 있으며, $w_{c_j c_j}$ 와 같은 반복적인 재귀도 있을 수 있다. t 시간에서, c_j 의 출력 y^{c_j} 는 다음과 같이 계산된다.

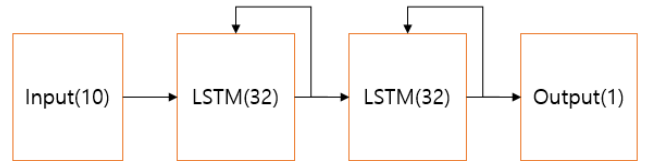
$$y^{c_j}(t) = y^{out_j}(t)h(s_{c_j}(t))$$

신경망의 상태 $s_{c_j}(t)$ 는 다음과 같이 나타낼 수 있다.

$$s_{c_j}(0) = 0$$

$$s_{c_j}(t) = s_{c_j}(t-1) + y^{in_j}(t)g(net_{c_j}(t)) ; t > 0$$

본 논문에서는 이 구조를 다음(그림 2)와 같이 구성하였다.



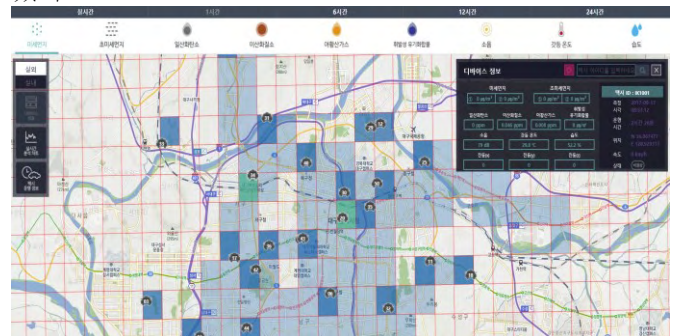
(그림 2) LSTM 구성도

본 논문에서 LSTM 모델은 이와 같이 2 개의 히든 레이어로 구성하였으며 10 개의 속성값을 입력 받고 1 개의 결과값(PM10)을 출력한다. Epoch 은 15 회로 Batch size 는 200 으로 설정하였다. MLP 모델도 마찬가지로 Epoch 15 회, Batch size 는 200 으로 설정하였으며, 히든 레이어는 3 개로 구성하였다.

4. 실험 및 평가

4.1 데이터

본 논문에서는 미세먼지 농도 예측에 활용하기 위해 택시의 갯등에 설치된 센서에서 수집되는 데이터 중 PM10, PM2.5, No2, Voc, So2 농도 수치를 사용하였다. 이 데이터는 택시 운행 시 10 초마다 센싱 하여 데이터를 저장한다. KISTI 에서는 다음 (그림 3)과 같이 데이터를 실시간으로 http://220.123.184.109:8080/KISTI_Web/에서 제공하고 있다.



(그림 3) 이동형 도시 환경 센싱 테스트 베드

이와 함께 추가적으로 기상청의 외부 분 단위 정보(온도, 습도, 풍향), 비산먼지 발생 사업장과 공사장 데이터를 사용하였다. 택시에서 수집되는 온도와 습도의 경우 외부가 아닌 갯등 안의 온도와 습도를 측정하기 때문에 갯등 내부 환경에 영향을 많이 받아 기

상청에서 제공하는 최소 단위(분) 데이터를 사용했으며, 센서 gps 값을 기준으로 반경 1KM 이내의 사업장과 공사장 총 수를 각각 입력하였다. 완성된 데이터는 <표 2>와 같다.

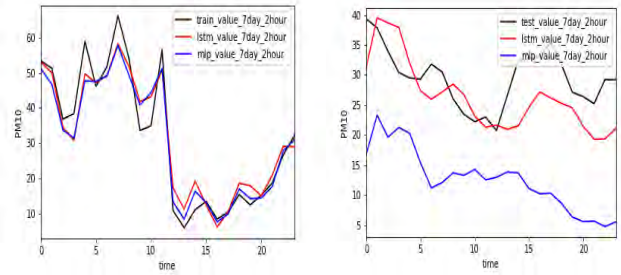
<표 2> 학습 데이터

TIME	PM2.5	PM10	NO2	VOC	SO2	LAT	LOX	TEMP	WIND SPEED	HUM	FACTORY	CONSTRUCTION
2017-07-01 00:00	24	24	0.012	18.5	0.004	35.831864	128.623095	26.2	2.1	78.2	0	1
2017-07-01 00:00	13	14	0.012	32.5	0.006	35.943973	128.566534	26.2	2.1	78.2	1	0
2017-07-01 00:00	55	62	0.013	6.3	0.004	35.8689723	128.632422	26.2	2.1	78.2	0	2
2017-07-01 00:00	18	20	0.013	35	0.004	35.865202	128.579657	26.2	2.1	78.2	0	2

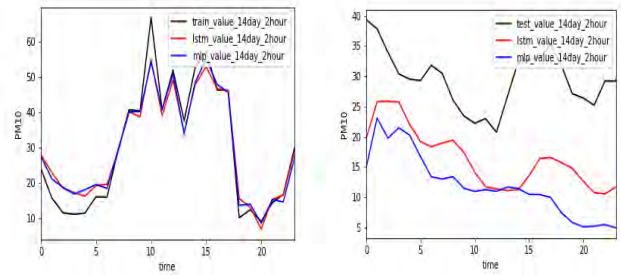
학습은 데이터를 1 시간 단위로 구성하여 실시되었으며, 시간대별로 데이터 개수의 편차가 큰 관계로 랜덤 추출을 하여 시간대별 개수가 똑같도록 조정하였다. 현재 시각이 k 일 때 t 시간 후의 미세먼지 농도를 예측하기 위해 MLP 와 LSTM 둘 다 정답을 y_{k+t} 로 설정하였다. 입력 데이터는 MLP 에서는 x_k 로 구성하였고, LSTM 에서는 $(x_k, x_{k+1}, x_{k+2}, \dots, x_{k+t-1})$ 형태로 구성하였다.

4.2 실험결과

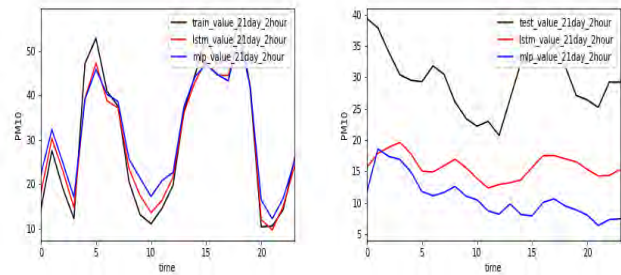
본 논문에서는 K 일의 t 시간 후 미세먼지 농도를 예측하기 위해 (K-1, K-2, ..., K-n)일의 데이터를 사용하였으며 이번 실험에서는 한달 치의 데이터를 학습시켰다. 우선 학습에 적합한 기간을 찾기 위해 t 값을 2 로 고정 하고 학습을 시도하였다. 다음 그림들은 학습 기간 변경에 따른 그래프와 평균 제곱근 오차 (Root Mean Squared Error, RMSE) 값 그리고 RMSE 변화 그래프를 나타냈다.



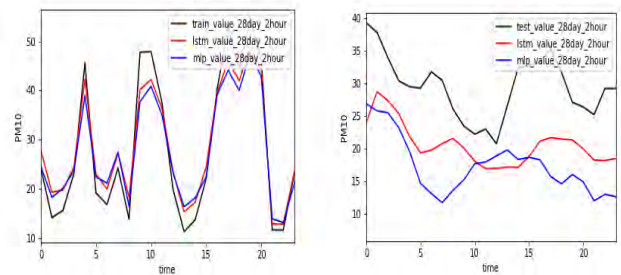
(그림 6) 7 일 데이터 학습 및 검증
 Training RMSE LSTM:18.40 MLP:18.06
 Test RMSE LSTM:18.25 MLP:24.02



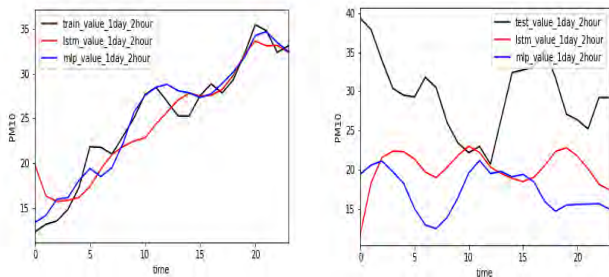
(그림 7) 14 일 데이터 학습 및 검증
 Training RMSE LSTM:17.67 MLP:17.69
 Test RMSE LSTM:21.40 MLP:24.48



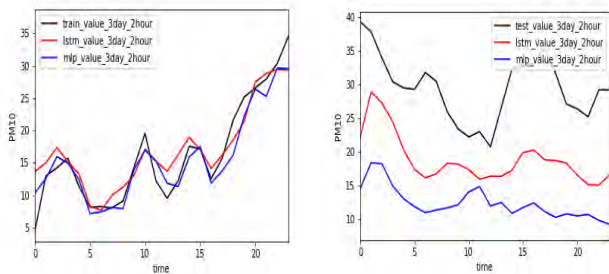
(그림 8) 21 일 데이터 학습 및 검증
 Training RMSE LSTM:17.10 MLP:17.74
 Test RMSE LSTM:21.51 MLP:25.15



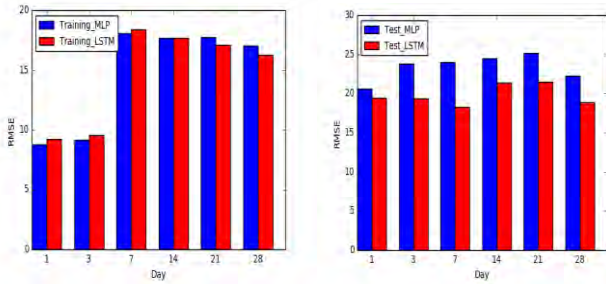
(그림 9) 28 일 데이터 학습 및 검증
 Training RMSE LSTM:16.29 MLP:17.01
 Test RMSE LSTM:18.90 MLP:22.27



(그림 4) 1 일 데이터 학습 및 검증
 Training RMSE LSTM:9.22 MLP:8.79
 Test RMSE LSTM:19.44 MLP:20.60



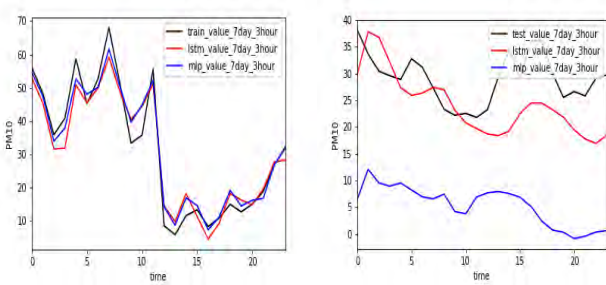
(그림 5) 3 일 데이터 학습 및 검증
 Training RMSE LSTM:9.57 MLP:9.13
 Test RMSE LSTM:19.93 MLP:23.81



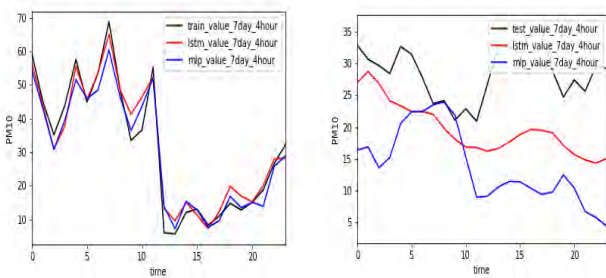
(그림 10) 학습 및 검증 RMSE 변화

학습 시에는 짧은 기간의 데이터(1 일, 3 일)를 학습시켰을 때의 RMSE 값이 제일 적었으나, 검증 시에는 7 일의 데이터를 학습한 것이 검증 데이터의 흐름을 잘 따라갔고 값에 가장 근접했다. MLP의 경우 데이터의 흐름을 따라가려는 경향은 보였으나 값에 근접하지는 못했다.

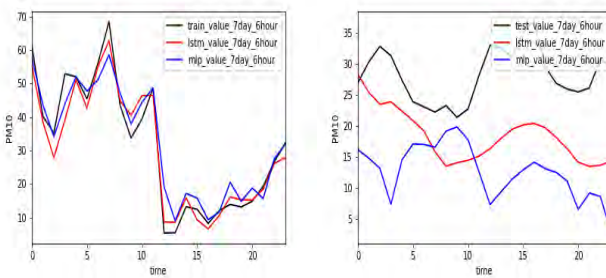
이 결과를 이용하여 학습 기간을 학습 결과가 가장 좋았던 7 일로 고정하고 t의 값에 변화를 주었다. 다음 그림들은 시간 변화에 따른 그래프와 RMSE의 값을 나타냈다.



(그림 11) 3 시간 후 데이터 학습 및 검증
 Training RMSE LSTM:18.16 MLP:17.79
 Test RMSE LSTM:18.51 MLP:29.05



(그림 12) 4 시간 후 데이터 학습 및 검증
 Training RMSE LSTM:17.83 MLP:17.95
 Test RMSE LSTM:18.88 MLP:23.10



(그림 13) 6 시간 후 데이터 학습 및 검증
 Training RMSE LSTM:17.83 MLP:18.71
 Test RMSE LSTM:19.13 MLP:23.48

LSTM의 경우 RMSE에서는 큰 변화가 없으나 2시간과 3시간 후의 예측 결과가 가장 좋았으며 t의 값이 커질수록 검증 데이터의 흐름을 따라가지 못하는 경향을 보였다. MLP 또한 t 값이 커질수록 그래프 상의 오차가 점점 더 커지는 모습을 확인할 수 있었다.

5. 결론

본 논문은 효율적인 도로 청소차량 운행 경로 제안을 위해 고해상도 지역별 오염 데이터와 지역의 특성을 고려하여 미세먼지 농도 예측에 활용하였다. 실험을 위해 2017.6~2017.7 약 두 달 동안 수집된 IoT 데이터와 외부 대기(온도, 습도, 풍속) 데이터, 비산먼지 발생 사업장 데이터를 통합하여 하루 4800 개의 데이터를 실험 데이터로 이용하였다. 데이터 학습에는 딥러닝 알고리즘 중 하나인 LSTM을 사용하였고 성능 비교를 위해 MLP를 이용하였다. 이번 연구에서는 2개월 분량의 테스트베드 데이터를 사용, 계절적 요인을 고려하지 못했고 지역별로 세분화 하기에는 데이터 양이 많지 않아 지역 전체를 데이터 셋으로 만들고 학습하였다. 실험결과 시간 속성을 고려한 LSTM이 MLP보다 평균 제곱근 오차 값이 경우에 따라 최대 30% 더 작음을 확인할 수 있었다. 향후 많은 양의 데이터가 축적되었을 때 지역을 세분화 한 데이터로 좀 더 가까운 의미의 고해상도 데이터 분석을 할 수 있을 것으로 예상되며, 교통량 등의 데이터를 추가해 더 안정적이고 정확한 예측값을 도출해 낼 수 있을 것으로 기대된다. 본 연구를 기반으로 고해상도 사물 데이터 기반 예측 연구의 가능성을 보였으며, 미세먼지 예측 결과를 활용 유연하고 효과적인 도로 청소차량의 운행 경로를 설정에 활용될 수 있을 것으로 기대한다.

참고문헌

- [1] 국립환경과학원(2015), '2013 국가 대기오염물질 배출량'
- [2] 수원대학교 강헌 외 6명(2012), '타이어 및 브레이크 패드 마모에 의한 비산먼지 배출량 및 위해성 조사'
- [3] 서울대학교 구윤모 외 2명(2017), '미세먼지 저감 대책의 비용효과분석'
- [4] 인천대학교 이희관 외 6명(2008), '도로 재비산 먼지 저감 시범사업 타당성 조사 연구'
- [5] 환경부(2016), '도로청소 매뉴얼'
- [6] 서강대학교 정보통신대학원 신혜영(2017), '미세먼지 농도 예측을 위한 통계 및 심층신경망 예측 기법 비교'
- [7] 한림대학교 오병두 외 2명(2016), 'Machine-Learning을 활용한 미세먼지(PM10) 농도 예측'
- [8] Sepp Horchreiter 외 1명(1997), 'LONG SHORT-TERM MEMORY'