

LSTM을 이용한 PM10 미세먼지 농도 예측

PM10 Particulate Matters Concentration Prediction using LSTM

저자 (Authors)	조경우, 정용진, 이종성, 오창현 Kyoung-woo Cho, Yong-jin Jung, Jong-sung Lee, Chang-heon Oh
출처 (Source)	한국정보통신학회 종합학술대회 논문집 23(2) , 2019.10, 632-634(3 pages)
발행처 (Publisher)	한국정보통신학회 The Korea Institute of Information and Communication Engineering
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09262546
APA Style	조경우, 정용진, 이종성, 오창현 (2019). LSTM을 이용한 PM10 미세먼지 농도 예측. 한국정보통신학회 종합학술대회 논문집, 23(2), 632-634
이용정보 (Accessed)	부산도서관 210.103.83.*** 2021/09/24 14:01 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

LSTM을 이용한 PM10 미세먼지 농도 예측

조경우 · 정용진 · 이종성 · 오창현*

한국기술교육대학교

PM10 Particulate Matters Concentration Prediction using LSTM

Kyoung-woo Cho · Yong-jin Jung · Jong-sung Lee · Chang-heon Oh*

Korea University of Technology and Education

E-mail : pinokio622@koreatech.ac.kr

요 약

산업화로 인한 대기오염 현상의 증대로 미세먼지에 의한 인체 영향에 관해 관심이 높아지고 있다. 이로 인해 미세먼지 예보의 중요성이 높아지고 있으나, 랜덤 특성을 갖는 미세먼지 농도 변화로 인해 예측 정확도 향상에 어려움이 있다. 본 논문에서는 PM10 미세먼지 농도 예측을 위해 딥러닝 기법 중 LSTM을 이용한 미세먼지 농도 예측을 수행한다. 이를 위해 기상 인자 및 대기오염 인자를 활용하여 예측 모델을 설계하고 예측 농도값 및 4단계의 미세먼지 AQI 예측 정확도를 비교한다.

ABSTRACT

Due to the increase of air pollution caused by industrialization, attention is increasing about the human effect by particulate matters. As a result, the importance of the particulate matters prediction is increasing, but there is a difficulty in improving the prediction accuracy due to the change of the particulate matters concentration having random characteristics. In this paper, we estimate the particulate matters concentration using LSTM in deep learning method for PM10 particulate matters concentration prediction. For this purpose, a predictive model is designed using weather and air pollution factors, and the predicted concentration values and four levels of fine dust AQI prediction accuracy are compared.

키워드

Particulate matters, Prediction, Deep learning, Recurrent neural network, Long short-term memory

I. 서 론

산업화로 인한 대기오염 현상의 증대로 미세먼지 문제에 관한 관심이 높아지고 있다. 또한, 미세먼지가 원인이 되는 다양한 인체 영향에 대한 발표로 인해 선진 주요 도시 대비 높은 대기오염 수준을 나타내는 국내 환경에서 미세먼지 예보의 중요성이 높아지고 있다[1]. 이로 인해 미세먼지 예보 정확성을 높이려는 노력이 진행되고 있으나, 랜덤한 특성을 갖는 미세먼지 농도 변화로 인해 기계 학습 기법을 사용하여 농도 예측 정확도를 향상시키려는 다양한 시도가 이루어지고 있다[2-5]. 본 논문에서는 PM10 미세먼지 농도 예측을 위해 딥러닝 기법의 하나인 LSTM(Long Short-Term Memory)을 이용한 미세먼지 예측을 수행한다. 이를 위해 기상 인자 및 대기오염 인자를 활용하여 예측 모델을 설계하고 예측 농도값 및 4단계의 미세먼지 AQI(Air Quality Index) 예측 정확도를 비교한다.

II. LSTM 기반 미세먼지 예측 모델

2.1 LSTM

예측 모델에 사용된 LSTM은 순환 신경망(Recurrent Neural Networks; RNN)의 한 종류로서

* corresponding author

순환 신경망이 긴 시퀀스를 학습하는 데 있어 step 이 진행됨에 따라 이전의 timestep 정보를 학습하지 못하는 기울기 소실 문제(vanishing gradient problem)를 극복한 모델이다. 기존 RNN과 달리 순환 신경망 내에 있는 메모리들을 input, forget, output gate를 포함하고 있는 cell로 대체하여 주어진 데이터의 흐름 조절을 통해 장기간의 연관성을 학습할 수 있는 장점이 있다[6],[7]. 그림 1은 LSTM의 cell 구조를 나타낸 그림이다. 이때, h 와 c 는 각각 t 시점에서의 hidden state, cell state를 의미한다.

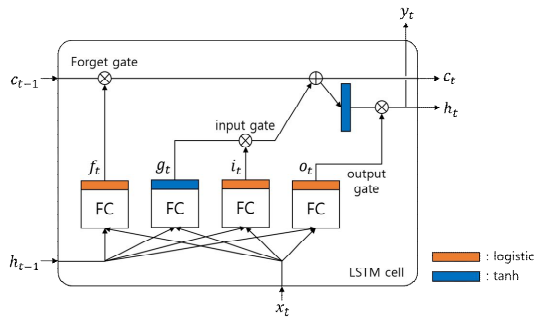


그림 1. LSTM cell 구조

2.2 데이터 구성 및 예측 모델

예측 모델에 사용된 데이터의 경우, 과거 10년간 (2009~2018년) 천안 지역의 온도, 평균 풍속, 최대 풍향, 습도와 같은 시간별 기상 데이터와 O_3 , NO_2 , CO , SO_2 , PM_{10} 과 같은 대기오염 물질 데이터를 활용하였다. 각 데이터는 기상자료개방포털과 에어코리아의 최종 확정 자료를 활용하였으며, 대기오염 물질 데이터의 경우 측정소 장비 유지 보수와 같은 결측 데이터를 최소화하기 위해 천안 지역 측정소 3곳의 데이터를 평균하여 활용하였다 [5]. 이후 수집된 전체 데이터를 75%의 training set, 25%의 test set으로 사용하였다. 또한, 각 dataset을 LSTM 모델의 input으로 사용할 수 있도록 범주형 데이터인 최대 풍향을 one-hot encoding 후 min max scaling을 거친 뒤 24시간의 sequence data로 재구성하였다.

모델 구조의 경우, 1개의 LSTM layer와 완전 연결 출력층으로 구성하였으며, LSTM layer와 출력층 사이에 dropout 층을 두어 모델을 설계하였다. 이때, input shape는 24시간의 sequence data 입력을 위해 x 특성 24개, timesteps 24로 설정하였으며, 모델의 최적화 함수는 adam을 사용하였다. 이후 해당 모델 구조의 최적 hyperparameter를 선정하기 위해 keras-tuner library를 사용하여 randomsearch 작업을 수행하였다. 표 1은 선정된 hyperparameter를 나타낸다.

표 1. 예측 모델 hyperparameter

Parameter	Value
units	20
dropout rate	0.3
batch size	40

III. 실험 결과

본 논문에서 설계된 LSTM 기반 미세먼지 예측 모델의 성능 평가를 위해 2장에서 선정된 hyperparameter를 사용하여 모델을 학습하였다. 모델 훈련 과정에서 training set의 80%는 train data, 20%는 validation data로 활용하였으며, 과다학습으로 인한 모델의 과대 적합을 막기 위해 early stopping callback을 사용하여 20회 이상 검증 손실이 향상되지 않을 때 학습을 조기 종료하였다. 그림 2는 test set 중 2018년 10월 1일부터 3개월 간의 모델 예측값과 실제 미세먼지 농도값을 비교한 그래프이다.

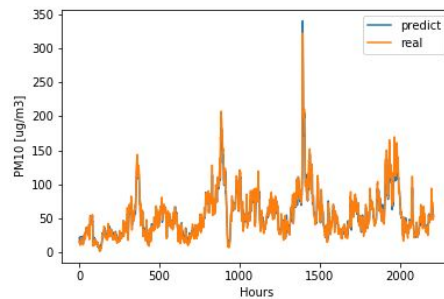


그림 2. 3개월 미세먼지 농도 예측 결과

예측 그래프 확인 시, 실제 미세먼지 농도값과 예측값의 추세가 근사한 모습을 보였다. 또한, $300 \mu g/m^3$ 이상의 미세먼지 농도에 대해선 모델의 예측값이 과대 예측된 모습을 볼 수 있으나, 해당 시점 외의 예측값은 실제 농도와 근사한 모습을 보였다. 표 2는 모델 학습 과정에서 validation data의 RMSE와 전체 test set의 AQI 별 예측 정확도를 나타낸다.

표 2. 예측 모델 RMSE 및 AQI 예측 정확도

	실제 빈도	예측 정확도
RMSE ($\mu g/m^3$)	-	10.21
total	21,781	87.59%
좋음	6,563	80.94%
보통	13,127	93.01%
나쁨	1,955	75.09%
매우 나쁨	136	65.44%

예측 모델의 전체 AQI 예측 정확도는 87%로 나타났다. $80\mu g/m^3$ 이하의 저농도 미세먼지에 비해 고농도 미세먼지의 예측 성능이 상대적으로 저조한 모습을 보였다. AQI 지수를 기준으로 고농도의 예측 오류를 확인한 결과, 평균 $5.55\mu g/m^3$ 의 예측값 차이를 보여 예측값과 실제값과의 농도 차이가 크지 않음을 확인하였다.

IV. 결 론

본 논문에서는 PM10 미세먼지 농도 예측을 위해 딥러닝 기법의 하나인 LSTM을 이용한 미세먼지 예측을 수행하였다. 이를 위해 천안 지역의 10년간 기상 및 대기오염 인자를 활용하여 예측 모델을 설계하고, 설계된 모델의 최적 hyperparameter를 사용하여 모델을 학습시켰다. 학습된 모델을 사용하여 미세먼지 농도를 예측한 결과, 전체 AQI 예측 정확도는 87%의 성능을 보였으며, 저농도 예측에 비해 상대적으로 고농도 예측 성능이 저조한 모습을 보였다. 그러나, AQI 지수를 기준으로 고농도의 예측 오류를 확인한 결과, 평균 $5.55\mu g/m^3$ 의 농도 차이를 보여 실제 미세먼지 농도와 근사한 농도값을 예측하는 것을 확인하였다. 향후 LSTM의 layer를 추가하여 모델의 layer에 따른 예측 성능 분석을 수행할 계획이다.

Acknowledgement

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2019R1I1A3A01059038).

References

- [1] Ministry of Science and ICT et. al., "Roadmap(plan) for the development of particulate matter technology," Policy Report, 2018.
- [2] A. Chaloulakou, G. Grivas, and N. Spyrellis, "Neural Network and multiple regression models for PM10 prediction in Athens: A comparative assessment," Journal of the Air & Waste Management Association, vol. 53, no. 10, pp. 1183-1190, 2003.
- [3] M. M. Dedovic, S. Avadakovic, I. Turkovic, N. Dautbasic, and T. Konjic, "Forecasting PM10 concentrations using neural networks and system for improving air quality," in proceeding of 2016 XI International Symposium on

- Telecommunications (BIHTEL), Sarajevo, pp. 1-6, 2016.
- [4] J. W. Cha, and J. Y. Kim, "Development of data mining algorithm for implementation of fine dust numerical prediction model," Journal of the Korea Institute of Information and Communication Engineering, vol. 22, no. 4, pp. 595-601, 2018.
- [5] K. W. Cho, J. S. Lee, and C. H. Oh, "Particulate Matter AQI Index Prediction using Multi-Layer Perceptron Network," in Proceeding Conference on Korea Information and Communication Engineering, Daejeon, vol. 23, no. 1, pp. 540-541, 2019. 05.
- [6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997. 11.
- [7] S. K. Kim and T. I. Oh, "Real-Time PM10 Concentration Prediction LSTM Model based on IoT Streaming Sensor data," Journal of Korea Academy Industrial Cooperation Society, vol. 19, no. 11, pp. 310-318, 2018. 11.