

GRU를 이용한 미세먼지(PM10) 농도 예측 모델

Particulate Matter(PM10) Concentration Prediction Model using GRU

저자 (Authors)	정용진, 조경우, 이종성, 오창현 Yong-jin Jung, Kyoung-woo Cho, Jong-sung Lee, Chang-heon Oh
출처 (Source)	한국정보통신학회 종합학술대회 논문집 23(2) , 2019.10, 644-646(3 pages)
발행처 (Publisher)	한국정보통신학회 The Korea Institute of Information and Communication Engineering
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09262550
APA Style	정용진, 조경우, 이종성, 오창현 (2019). GRU를 이용한 미세먼지(PM10) 농도 예측 모델. 한국정보통신학회 종합학술대회 논문집, 23(2), 644-646
이용정보 (Accessed)	한신대학교 118.131.***.99 2020/01/03 17:30 (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.

GRU를 이용한 미세먼지(PM10) 농도 예측 모델

정용진 · 조경우 · 이종성 · 오창현*

한국기술교육대학교

Particulate Matter(PM10) Concentration Prediction Model using GRU

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

Korea University of Technology and Education

E-mail : jungyj0211@koreatech.ac.kr

요 약

정확한 미세먼지 예보 서비스를 위해 최근 신경망 알고리즘을 이용한 미세먼지 농도 예측의 연구가 진행되고 있다. 본 논문에서는 시계열 특성을 가지는 미세먼지 농도의 예측에서 순환신경망 알고리즘 중 GRU 알고리즘을 이용하여 미세먼지 농도 예측 모델을 설계한다. 기상 데이터와 대기오염 물질 데이터를 활용하며, 최적의 hyper parameter를 적용하여 미세먼지 농도 예측을 수행한다. 또한 예측 모델의 성능 평가를 위해 AQI 기반의 예측 정확도 평가를 진행한다.

ABSTRACT

Recently, there are many researches on particulate matter concentration prediction using neural network algorithm for accurate particulate matter prediction service. In this paper, design the particulate matter concentration prediction model using GRU algorithm among the recurrent neural network algorithms for the prediction of particulate matter concentration with time series characteristics. It uses the weather data and air pollutant data and predicts the particulate matter concentration by applying the optimal hyper parameter. In addition, AQI based prediction accuracy is evaluated to evaluate the performance of the prediction model.

키워드

Particulate matter, Deep learning, Recurrent neural networks, GRU

I. 서 론

다양한 오염원으로 발생하는 미세먼지는 호흡기 질환, 심혈관 질환 등 건강에 영향을 주는 요인으로 위험성 물질로 분류되고 있으며, 이와 같은 영향으로 도시 경제 및 사회 활력 저하의 원인으로 분석되고 있다[1],[2]. 생활에 직접적인 영향을 주는 미세먼지에 대한 피해를 줄이고자 정부기관 및 민간 기업들은 미세먼지 감축 및 예측을 위한 연구를 다양한 방법으로 진행하고 있으며, 4단계로 구분된 AQI(Air Quality Index)를 기준으로 미세먼지 실시간 정보 제공 및 예보 서비스를 제공하고 있다.

본 논문에서는 미세먼지를 예측하기 위해 순환

신경망 알고리즘 중 하나인 GRU(Gated Recurrent Unit) 알고리즘을 이용하여 미세먼지를 예측한다 [3]. 이를 위해 기상데이터와 대기오염물질 데이터를 활용하여 예측 모델을 설계하고 미세먼지 수준을 4단계로 구분하는 AQI를 기반으로 예측 성능 평가를 진행한다.

II. GRU 기반 미세먼지 예측 모델

2.1 GRU

GRU 알고리즘은 그림 1과 같이 순환신경망의 장기 의존성 문제를 해결하기 위해 설계된 LSTM(Long Short-Term Memory) 알고리즘의 input, output, forget에 대한 3개의 게이트로 구성된 구조

* corresponding author

를 update, reset에 대한 2개의 게이트로 알고리즘을 구성하여 단순화한 알고리즘이다[4].

LSTM 알고리즘과 동일하게 게이트를 이용하여 데이터의 양을 조절하지만 forget, input 게이트가 결합된 update 게이트를 적용하고 reset 게이트를 hidden state 값에 바로 적용함으로써 LSTM 알고리즘의 장기 의존성 문제 해결 능력을 유지하며, 동일한 예측 성능과 보다 빠른 학습 속도의 이점을 갖는다[5].

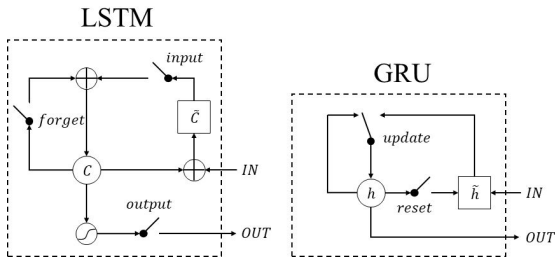


그림 1. LSTM, GRU cell gate 구조

2.2 데이터 구성 및 전처리

학습을 위한 데이터는 천안 지역의 2009~2018년 간 시간별로 수집된 기상데이터와 대기오염물질 데이터를 활용하였다. 기상데이터는 온도, 습도, 평균 풍속, 최대 풍향으로 구성되며, 대기오염물질 데이터는 O_3 , NO_2 , CO , SO_2 , PM_{10} 으로 구성하였다. 최대 풍향의 경우 범주형 데이터이므로 one-hot encoding을 통해 16개의 수치형 데이터로 변환하여 사용하였다. 이후 전체 데이터를 min max scaling을 적용하여 동일한 데이터 크기로 변환하였으며, 학습 및 테스트를 위해 training set(60%), validation set(15%), test set(25%)으로 dataset을 구성하였다[6].

2.3 예측 모델 설계

모델의 경우 GRU cell로 구성된 1개의 hidden layer와 출력을 위한 1개의 dense layer로 구성하였으며, 각 층 사이에 over fitting을 방지하기 위한 dropout layer를 구성하여 모델을 설계하였다. GRU 알고리즘의 경우 순환을 위해 input shape가 sequence data 형태로 구성되어야 한다. timesteps 설정을 통해 input shape를 구성할 수 있으며, 24시간을 기준으로 sequence data 형태를 구성하기 위해 timesteps를 24로 설정하였다. 구성된 모델의 효율적인 학습을 위한 최적화 함수는 adam으로 이용하였다. 학습의 양을 의미하는 batch size는 100으로 설정하며, 학습의 반복 횟수를 의미하는 epoch의 경우 200으로 설정한 뒤 early stopping을 적용하였다. Early stopping은 훈련과 검증으로 구분되는 학습 과정에서 훈련 시 발생하는 손실 값과 검증 시 발생하는 손실 값의 차이가 일정 횟수 이상 차이가 없을 경우 조기 종료함으로써 모델의 과도한 학습으로 인해 발생할 수 있는 over fitting을 방

지하는 역할을 한다. 이후 모델 구성 요소 중 hidden layer를 구성하는 neuron의 수와 dropout의 최적 값을 찾기 위해 keras tuner library를 이용하여 hyper parameter 탐색 작업을 수행하였다. 작업을 통한 상위 3위의 최적 값은 표 1과 같다.

표 1. Hyper parameter top3

Rank	units	Dropout	Epoch
1	40	0.1	125
2	40	0.2	147
3	60	0.3	107

III. 실험 결과 및 평가

GRU 기반의 미세먼지 예측 모델 성능 평가를 위해 표 1의 1위 hyper parameter를 적용하여 2장에서 구성한 train set과 validation set을 활용하여 모델 학습을 진행하였다. 이후 학습이 완료된 모델을 통해 test set을 이용한 미세먼지 농도 예측을 진행하였다. 그림 2는 예측 모델을 통해 예측된 미세먼지 농도 값과 실제 값을 비교한 그래프이다.

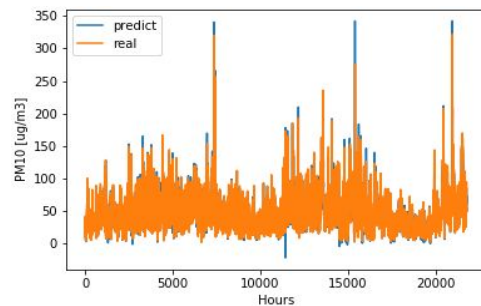


그림 2. 미세먼지 농도 예측 결과

실제 값과 예측 값의 경우 유사한 추세를 보이거나 미세먼지 농도 $250\mu g/m^3$ 이상의 구간에서 $300\mu g/m^3$ 이상의 값을 예측하는 것을 통해 해당 구간에서의 농도값의 과대 예측 결과를 확인할 수 있었다. 또한 예측 값과 실제 값의 평균 오차를 확인한 결과 $5.5759\mu g/m^3$ 의 차이를 보였으며, 전반적으로 실제 값과 예측 값의 농도 차이가 크지 않음을 확인하였다.

표 2. AQI 구분 예측 정확도

구분	실제 빈도	일치 개수	정확도
좋음	6,563	5,485	83.57%
보통	13,126	11,993	91.36%
나쁨	1,955	1,495	76.47%
매우 나쁨	136	98	72.05%
total	21,780	19,071	87.56%

표 2는 AQI 기반으로 구분된 미세먼지 농도의 좋음, 보통, 나쁨, 매우 나쁨의 예측 정확도이다. 전체 정확도는 87.56%를 보이나 나쁨과 매우 나쁨의 경우 76.47%와 72.05%의 저조한 수준의 정확도를 보였다. 이는 수치형 데이터인 예측 값을 범주형으로 구분하여 정확도를 산출하는 과정에서 발생하는 문제일 가능성이 있으므로 $5\mu g/m^3$ 의 오차 범위를 적용하여 예측 정확도를 측정하였다.

표 3. 오차 범위($5\mu g/m^3$) 적용 예측 정확도

구분	실제 빈도	일치 개수	정확도
좋음	6,563	6,266	95.47%
보통	13,126	12,777	97.34%
나쁨	1,955	1,702	87.05%
매우 나쁨	136	106	77.94%

표 3은 오차 범위를 적용한 예측 정확도이다. 오차 범위를 적용한 결과 저농도에 해당하는 좋음과 보통의 경우 각 11.9%, 5.98% 상승한 예측 정확도를 보였으며, 고농도에 해당하는 나쁨과 매우 나쁨의 경우 10.58%, 5.89% 상승한 예측 정확도를 보였다. 전반적으로 상승한 예측 정확도를 보였으나 매우 나쁨은 미비한 예측 정확도의 상승을 보였다.

IV. 결론

본 논문에서는 미세먼지 농도(PM10) 예측을 위해 순환신경망 기반의 GRU 알고리즘을 이용한 모델 설계 및 예측을 수행하였다. 학습 및 평가를 위한 데이터는 천안 지역의 10년간 기상 및 대기오염물질 데이터를 활용하였다. keras tuner library를 통해 hidden layer의 unit 수와 dropout의 최적 값을 도출하여 모델에 적용 후 학습하였다. 학습된 모델을 통해 미세먼지 농도를 예측한 결과 실제 값과 비교하여 미세먼지 농도 $250\mu g/m^3$ 이상의 구간에서 과대 예측 결과를 확인할 수 있었다. 전반적인 예측 값의 평균 오차의 경우 $5.5759\mu g/m^3$ 의 값으로 실제 값과 차이가 크지 않음을 확인하였다. AQI 기준으로 측정한 예측 정확도는 87.56%를 보였으나 나쁨과 매우 나쁨의 경우 70% 수준의 예측 성능을 보였다. 이는 범주형으로 구분하여 정확도를 측정하는 과정에서 발생하는 문제라 볼 수 있으며, 오차 범위를 적용한 결과 전반적인 예측 정확도가 상승한 것을 확인할 수 있었다. 저농도의 경우 95% 이상의 예측 정확도를 보인 반면 고농도의 경우 90% 이하의 예측 정확도를 보였다. 특히 매우 나쁨의 경우 예측 정확도의 상승폭이 미비하며, 이는 매우 나쁨의 학습을 위한 표본의 수가 적어 발생하는 문제로 볼 수 있다. 향후 매우 나쁨의 정확도 향상을 위해 적은 표본에서도 뚜렷한 효과

를 주기위한 모델 설계를 진행하여 예측 모델의 성능 향상을 수행할 계획이다.

Acknowledgement

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

References

- [1] G. W. Evans, "Air Pollution and Human Behavior," Journal of Social Issues, Vol. 37, No. 1, pp. 95-125, 1981.
- [2] M. S. Seo, "The Impact of Particulate Matter on Economic Activity," The Korean Women Economists Association, Vol. 12, No. 1, pp. 75-100, Jun. 2015.
- [3] Tensorflow API Document [Internet]. Available : https://www.tensorflow.org/api_docs/python/tf
- [4] D. G. Lee, Y. G. Sun, I. S. Sim, Y. M. Hwang, S. W. Kim, and J. Y. Kim, "Prediction of Power Consumptions Based on Gated Recurrent Unit for Internet of Energy," The Journal of Institute of Korean Electrical and Electronics Engineers, Vol. 23, No. 1, pp. 120-126, Mar. 2019.
- [5] K. H. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder - Decoder for Statistical Machine Translation," arXiv preprint arXiv:1406.1078, Jun. 2014.
- [6] 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, May. 2019.