Journal of Korea Multimedia Society Vol. 26, No. 8, August 2023(pp. 911-924)
https://doi.org/10.9717/kmms.2023.26.8.911

미세먼지 농도 예측을 위한 어텐션 기반 합성곱 장단기 메모리 모델

이준민†, 김경태†, 최재영†

Attention Based-ConvLSTM-DNN Networks for Fine Dust Concentration Prediction

Joon Min Lee†, Kyeong Tae Kim†, Jae Young Choi†

ABSTRACT

Air pollution, particularly fine dust, poses a significant threat to public health and necessitates accurate prediction models for effective mitigation strategies. In this paper, we propose a so-called attention-based ConvLSTM-DNN networks for fine dust concentration prediction. Our proposed model integrates the feature extraction capabilities of a 2D Convolutional Neural Network (CNN) with the long-term memory retention of an LSTM, capturing spatial and temporal dependencies in the input data. We apply an attention mechanism to enhance the model's predictive ability by assigning greater weight to features of importance. Using real-world weather and fine dust concentration data, we conduct experiments to evaluate our model. We demonstrated the superior prediction accuracy of our model through comparative analyses with the state-of-the-art (SOTA) CNN-LSTM model. We evaluated prediction performance of our model using RMSE, ACC, POD metrics. Comparative analysis with existing models revealed that our proposed model showed improvements in performance with a decrease of 2.89 in RMSE, an increase of 8.14% in ACC, and an increase of 7.12% in POD. The proposed model offers accurate and reliable fine dust forecasts, aiding decision-making processes. By leveraging ConvLSTM, attention mechanisms, and optimized pre-processing, our model can improve fine dust prediction, leading to advances of air pollution monitoring and control.

Key words: ConvLSTM, Deep Neural Network, Attention Mechanism, PM2.5, Air Quality

1. 서 론

미세먼지는 건강과 환경에 해로운 영향을 미치는 대기 오염물질 중 하나이다. 따라서 정확한 미세먼지 예측은 인간의 건강과 안전을 위해 매우 중요하다. 최근에는 미세먼지 입자 PM(particulate matter)의 높은 수치가 전 세계적인 문제로 주목받고 있다. 특히 PM_{10}과 $PM_{2.5}$는각각 $10(\mu m/m^3)$와 $2.5(\mu m/m^3)$의 크기를 가지는 공기 중 작은 미세 입자로 정의된다. 이러한 높은 농도의 PM이 인간에게 매우 유독하다

※ Corresponding Author : Jae Young Choi, Address: (우편번호) Engineering Dep. 423, 81, Oedae-ro, Mohyeon-eup, Cheoin-gu, Yongin-si, Gyeonggi-do, 17035, Korea, TEL : +82-10-4304-7422, E-mail : jychoi@hufs.ac.kr
Receipt date : Jul. 19, 2023, Approval date : Aug. 8, 2023
† Pattern Recognition and Machine Intelligence (PMI) Lab., Division of Computer Eng., Hankuk University of Foreign Studies
(Joon Min Lee, E-mail : joonminlee@hufs.ac.kr)
(Kyeong Tae Kim, E-mail : rudxo9251@gmail.com)

※ This research was supported by Hankuk University of Foreign Studies Research Fund.
※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1A2C1092322).
※ This research was supported by the "Development of technology using analysis of ocean satellite images (20210046)" funded by the Korea Institute of Marine Science & Technology Promotion (KIMST).

고 보여주고 있다[1]. 이전 연구들은 주로 기상 데이터와 미세먼지 농도 데이터를 사용하여 예측 모델을 개발했지만, 이러한 모델들은 미세먼지 농도 값만 학습하는 경향이 있어 부정확한 예측을 하는 경우가 많다. 본 논문에서는 딥러닝 모델을 사용하여 미세먼지 농도 데이터를 2D 그리드(Grid)로 만들어 이미지화 시키고, 지역별 데이터를 통해 더욱 정확한 미세먼지 예측 모델을 제안한다. 현행 환경부에서 운영중인 미세먼지 예보모델 CMAQ[2]의 적중률은 70%, 통합대기환경지수(CAI, Comprehensive air-quality index)[3] 기준 나쁨 이상의 고농도 감지 확률은 낮다. 기존의 수치 예보 모델은 배출량 및 기상자료 등의 입력 자료 불확실성과 복잡한 대기현상을 완전하게 반영하지 못하는 근본적인 한계를 지니고 있다. 실시간으로 수집되는 측정 및 예보자료의 빅데이터화, 데이터를 체계적인 처리와 분석을 위한 신경망 활용 모델 RNN(Recurrent Neural Network)[4], LSTM(Long Short-Term Memory)[5,6] 이 개발되었다. 이미지를 입력으로 사용하는 VGG16[7]을 사용하거나, CNN 네트워크를 사용한 미세먼지 예측 모델 [8]은 학습 파라미터가 많고, 예측 시간대(수행하는 작업에 따라 구분, 본 연구에서는 6시간 단위)별로 미세먼지 농도를 학습 및 예측하여 소요되는 시간이 길어 실시간 운용에 적합하지 않다. 이런 문제점을 극복하기 위해 제안 모델은 시간대별 예측을 하는 것이 아닌, 한 번의 학습으로 모든 예측 시간대를 예측하는 어텐션 기반 ConvLSTM-DNN 모델을 제안한다.

미세먼지 예측에 ConvLSTM[9]를 사용, DNN과 결합하여 미세먼지 농도 예측 네트워크를 설계하였으며, CBAM(Convolutional Block Attention Module)[10]의 Channel Attention 메커니즘을 적용하였다. ConvLSTM은 LSTM 구조에 2D(Dimension) 합성곱(Convolution) 연산을 추가한 네트워크로, 2D 입력 데이터의 공간적 패턴을 학습하는 것에 효과적이며, 이미지나 비디오와 같은 데이터 처리에 주로 사용된다. 제안하는 모델은, 입력으로 주어지는 공간 이미지 데이터의 패턴을 분석하고 학습하여, 1D 합성곱을 진행, 총 19개의 지역별 데이터에 따라 예측 시간대, 지역별로 미세먼지 농도를 예측한다. 복잡하고 무거운 예측 모델에 비해 간결한 예측 모델이며, 쓰임새에 맞게 전처리 된 지역별 기상 데이터에 대한 학습을 강화하고, 지역별 세부적인 예측을 위해 고안한 네트워크다. 어텐션 메커니즘은 입력 시퀀스에서 중요한 부분에 더 집중할 수 있도록 하는 기법으로, 모델이 입력 시퀀스에서 가장 중요한 부분에 더 집중하고 불필요한 부분은 억제해, CNN 특징(feature)에 어텐션 가중치(attention weight)를 적용해 특징에 대한 품질을 향상하고 예측성능을 개선한다.

CBAM은 컨볼루션 신경망(Convolution Neural Network)에 어텐션 기법을 적용하는 방법의 하나로 두 가지 종류의 어텐션을 사용한다. 채널 어텐션(Channel Attention)은 채널별로 중요한 정보를 추출하여 스케일링(Scaling)한다. 채널 어텐션을 통해 입력 이미지에서 각 채널의 정보 중 가장 중요한 정보를 추출한다. 공간 어텐션(Spatial Attention)은 입력 데이터의 공간적 패턴 중, 가장 중요한 부분을 강조한다. CBAM은 두 가지 어텐션을 조합하여 특징 맵(Feature map)의 각 위치에 대한 중요도를 계산하고, 이를 통해 입력 이미지의 중요한 부분에 더 집중할 수 있도록 한다. 모델은 전역 평균/최대 풀링(Global Average/Max Pooling)을 통해 입력 이미지의 중요한 부분을 분석하여 가중치를 추출하고, 기존 입력 데이터에 어텐션 가중치를 추가하여 불필요한 부분을 억제하고 중요한 부위에 모델이 더 집중하여 학습할 수 있도록 도와준다. 제안하는 모델에서는 채널 어텐션을 사용하였으며, 전통적인 기상 데이터와 배출량 계산 기반 수치 모델 CMAQ, 기존 예측 모델 CNN-LSTM과 비교했을 때, 예측 정확도, 탐지율, 오경보율이 제안하는 모델이 모든 부분에서 더 나은 예측 성능을 보여주었다.

2. 관련 연구

현재까지도 미세먼지 예측 연구에는 다양한 방법과 모델이 사용되고 있다. 기상 데이터와 대기 환경 데이터를 모두 사용하는 모델부터 기계 학습 및 인공신경망을 활용하는 모델까지 방법이 있다. 기상 데이터를 활용한 미세먼지 예측에 관한 연구는 매우 중요하다. 기상 요인은 미세먼지 농도에 큰 영향을 미치기 때문이다. 예를 들어, 기온, 습도, 바람, 일조량, 강수량 등의 기상 변수와 미세먼지 농도 간의 상관관계를 분석하고 이를 활용하여 미래 미세먼지 농도를 예측하는 방법[11,12]이 연구되었다. 랜덤포레스트

(Random Forest), 서포트벡터머신(Support Vector Mahchine) 및 인공신경망(Artificial Neural Network)을 활용, 수치 예측 모델과 기계 학습 알고리즘을 통합하여 서울 내 40개 관측소의 PM_{10}과 $PM_{2.5}$ 농도의 단기 예측 모델[13]을 제시한다. 결과는 머신러닝(Machine Learning) 기반 모델이 수치 기반 모델보다 높은 예측 정확도를 보였으며, 고농도 PM 수치도 정확도가 개선됐다.

CLSTM(Convolutional Long Short-Term Memory Network) 하이브리드 예측 모델[14] 을 기반으로 한 대기 오염 예측 시스템은 합성곱 신경망(CNN)과 장단기 메모리(LSTM) 네트워크를 결합하여 예측변수에서 중요한 특성을 자동으로 감지하고 추출하며, 다음 모델링 단계를 위한 시계열을 생성하는 모델이다. 제안된 모델은 경쟁 수치 모델인 LSTM과 비교하였을 때, 평균절대오차(Mean Absolute Error, MAE) 및 평균제곱근오차(Root Mean Squared Error, RMSE) 측면에서 우수한 성능을 보였다. CNN 및 LSTM 프레임워크의 퓨전이 예측 성과를 향상하며, CLSTM 하이브리드 프레임워크는 시간대별 데이터 예측에 상당히 강력하며 대기오염 모델 및 미래의 공중 보건 및 대기질 연구에 활용이 된다고 제안했다. 하지만 위 모델의 구조는 두 신경망 모델을 사용하는 네트워크는 신경망에서 다른 신경망으로 핵심 특징를 전달하는 과정에서 데이터의 손실을 야기 할 수 있다. LSTM 기반의 인코더(Encoder)-디코더(Decoder) 아키텍처를 기반으로 한, 다변량 시계열 데이터(Multivariate time series data)의 다중 시간대 예측을 위한 시퀀스-투-시퀀스(Sequence-to-Sequence) 딥러닝 모델[15]을 제안했다. 다변량 시계열 데이터의 지역적인 경향과 장기 의존성 패턴의 깊은 표현 특징을 학습하며, 대기질 예측 실험 결과, 제안된 모델은 기본적인 딥러닝 모델보다 더 좋은 성능을 보였으며, 온도, 습도, 풍속 및 PM_{10}과 $PM_{2.5}$와 같은 다변량 대기질 관련 시계열 데이터의 상호 의존성과 비선형 관계를 효과적으로 탐색하고 학습한다고 주장했다. 정규적이고 규칙적인 시계열 데이터에 대해서 좋은 성능을 보였지만 비정상적이거나 이상점이 포함된 시계열 데이터에 대해 대응하기 위한 후속 연구의 개선이 필요하다고 고찰했다. 위 관련 연구 논문과 같이 여러 기상 데이터를 입력으로 조합하거나, 변수 간의 상관관계에 중점을 맞추고, 미세먼지 농도 값에 영향을 주는 것을 모델이 인지할 수 있도록 설계하여, 최종적으로 미세먼지 농도 예측 시, 다른 기상 데이터가 예측 농도의 근거가 될 수 있도록 하는 방법 등이 제시되고 있다.

가정용 전력 소비량에 미치는 영향[16], 질병 간의 상관관계 분석[17], 학교시설 운영에 미치는 영향[18], 교통흐름에 미치는 영향[19] 등 미세먼지 데이터에 대한 실생활 밀착 연구가 활발하게 진행되고 있다. 그럼에도 불구하고 미세먼지 관련 데이터셋은 측정 위치 정보, 측정 일자 및 시간, 미세먼지 농도, 기상 정보 등 예측 모델을 개발하는 사람에 따라, 사용되는 용도가 상이하다. 측정소가 한정됨에 따라, 미세먼지 농도가 측정되지 않은 지역, 측정소 주변의 상세한 정보가 부족하여, 미세먼지 상황을 완벽하게 반영하기가 어렵다. 다양한 장비와 방법을 사용하여 측정하면, 같은 위치에서도 측정값의 불일치가 발생할 수 있으며, 노이즈(Noise)가 데이터의 신뢰성을 저하할 수 있다. 미세먼지 예측을 위해 고려해야 할 공간 및 시간적 상관관계를 활용하는 심층학습 프레임워크[20]는 단기간의 갑작스러운 기상 악화나, 측정이 제대로 되지 않는 경우 노이즈가 발생하여, 실질적인 예측성능이 저하 될 수 있으며, 날것의 측정 데이터를 바로 사용하는 것이 아닌 정제작업이 필요하다고 고찰하였다.

제안하는 모델은 6시간 단위로 예측이 되는 모델로, 6시간마다 측정된 데이터를 조합하여 모델에 입력하고 오늘(D+00), 내일(D+01), 모레(D+02)를 예측한다. 미세먼지 데이터셋은 일부 기간 또는 특정 지역에서 데이터가 부족하거나 누락된 경우가 있을 수 있다. 이를 해결하기 위해 보간법[21]을 사용하여 누락 된 데이터의 주변 값을 활용하도록 하였다. 본 연구는 어텐션 기반 ConvLSTM-DNN (Convolutional Long Short-Term Memory-Deep Neural Network) 네트워크를 활용한 미세먼지 예측에 대한 독창적인 접근법을 제시한다. 이 방법은 ConvLSTM과 DNN 아키텍처의 장점을 결합하면서 어텐션 메커니즘을 포함하여 예측 정확성을 향상시키고 미세먼지 데이터의 중요한 시공간 의존성을 포착한다. ConvLSTM 네트워크를 활용함으로써, 제안 모델은 시계열 데이터의 순차적인 특성을 효과적으로 다룰 수 있으며, 미세먼지 농도의 시간적 의존성과 패턴을 잘 포착하였다. ConvLSTM 레이어는 입력 데이터에

합성곱 연산을 수행하여 공간 정보를 보존하고 동시에 장기적인 의존성을 학습한다. 또한, DNN 레이어의 통합은 복잡한 비선형 관계를 학습하고 입력 데이터에서 고수준 특성을 추출하는 모델의 능력을 향상시킨다. DNN 레이어는 ConvLSTM 레이어에서 학습된 표현을 효과적으로 포착하고 처리함으로써 전체적인 예측력에 기여한다. 제안하는 접근법의 주요 독창성은 어텐션 메커니즘에 있다. 어텐션 기법을 적용함으로써, 예측 모델은 입력 데이터의 다른 공간 및 시간적 요소의 중요성을 동적으로 가중치를 부여하여 판단한다. 이를 통해 네트워크는 가장 관련성이 높고 정보를 가진 특징에 집중할 수 있다. 자세한 내용은 3장에서 설명한다.

3. 제안 미세먼지 농도 예측 방법

본 논문에서 제안하는 미세먼지 농도 예측을 위한 어텐션 기반 합성곱 장단기 메모리 모델의 전체적인 아키텍쳐는 Fig. 1과 같다. 시퀀스별로 나누어진 시계열 데이터를 제안한 ConvLSTM 네트워크에 입력하여 ConvLSTM 셀(cell)과 어텐션(Attention) 모듈을 통해 결과물로 각 시간대의 특징(Feature)을 얻게 되고, 추출된 특징을 1차원 형태로 만들어. 지역별 정보가 담긴 1차원 DNN data와 결합(Concatenate)하여. 완전 연결계층(FC Layer)을 통해 $T_{n+1} \sim T_{n+m}$의 농도를 예측한다. 본 연구에서는 실제값과의 비교는 평균제곱오차(Mean Squared Error, MSE)를 통해 학습하며, 최적화 함수는 ADAM(Adaptive Moment Estimation)을 사용하였으며, ConvLSTM과 DNN 네트워크에서 활성화 함수를 각각 ReLu(Rectified Linear Unit)와 PReLU(Parametric Rectified Linear Unit)를 사용하여 다층 구조에서 비선형성을 잘 반영하여, 중요한 정보의 추출과 전달을 효율적으로 진행하도록 설계하였다. 제안하는 모델의 모든 학습 과정은 End-to-End로 진행된다. 아래의 수식은 평균제곱오차로, n은 전체 데이터셋, y는 모델이 예측한 값, t는 실제 값이다.

$$MSE = \frac{1}{n}\sum_{i}^{n}(y_i - t_i)^2 \quad (1)$$

3.1 공간 이미지 데이터의 특징 추출

ConvLSTM은 합성곱 연산을 이용하여 입력과 상태 간의 연결을 만든다. 입력$(X_1,...,X_t)$과 은닉 상태(H_t)의 합성곱 연산$(*)$을 통해 입력 게이트(i_t)와 망각 게이트(f_t)를 계산하고, 기억 셀(C_t)의 이전 값(C_{t-1})과 새로운 입력을 이용하여 새로운 기억 셀 값을 계산한다. 출력 게이트와 기억 셀의 값을 이용하여 새로운 은닉 상태를 계산하며 보다 공간적인 정보를 더 잘 활용할 수 있게 된다. $(\circ)$은 아다마르 곱(Hadamard product)로 두 행렬의 각 성분을 곱하는 연산이다. 제안하는 네트워크는 공간 이미지 데이터

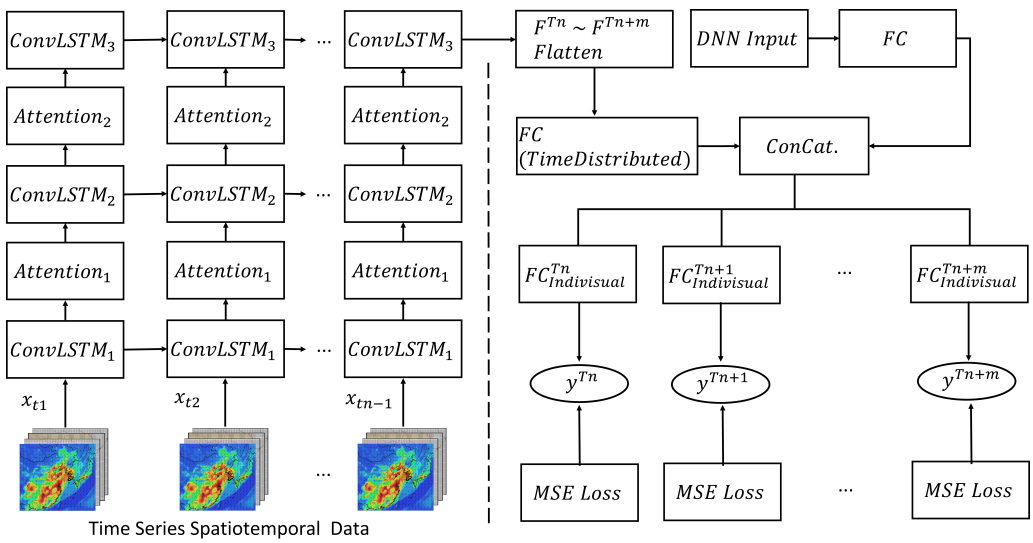

Fig. 1. Architecture of the proposed attention based ConvLSTM-DNN for fine dust concentration prediction.

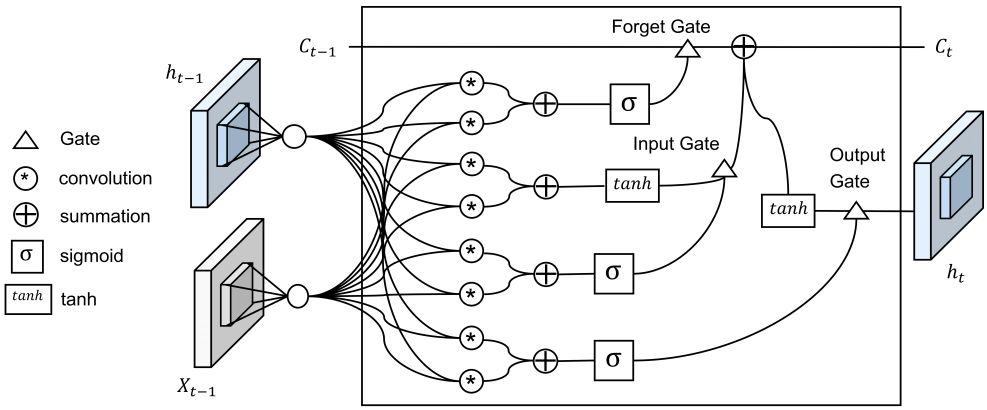

Fig. 2. The inner structure of a ConvLSTM cell [22].

로 (?, 12, 70, 90, 1)가 입력된다, 입력 데이터는 ConvLSTM2D 셀을 통해 특징 추출이 진행된다. 필터(filter) 개수는 (64), 커널(kernel)의 크기는 (5,5)로 구성되어있으며, 활성화 함수는 ReLu를 사용한다. ConvLSTM2D 셀을 통해 추출된 특징은 어텐션 모듈로 전달된다. 어텐션 모듈의 자세한 연산 과정은 3.2에서 설명 한다.

어텐션 모듈을 통해 가중치가 적용된 어텐션 특징 맵(Attention Feature map)은 ConvLSTM 연산과 결합하여 시퀀스의 각 타임 스텝에서 이전 타임 스텝의 은닉 상태와 셀 상태를 사용하여 다음 타임 스텝의 새로운 상태를 계산한다. 따라서, 입력 데이터의 공간적인 특징과 시간적인 의존성을 모두 포착한 정보를 담고 있다. 네트워크에는 기울기 소실(Gradient Vanishing) 및 안정화를 위하여 배치정규화(Batch Normalization)을 적용하였다. 마지막으로 미세먼지 농도 예측을 위한 예측 시간대(=12)로 나누어지며 DNN 레이어로 전달 및 지역별 데이터와 결합하기 위해 완전연결계층(FC-Layer)을 통해 1차원 데이터로 변환한다.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (3)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \quad (5)$$

$$H_t = o_t \circ \tanh(C_t) \quad (6)$$

3.2 채널 어텐션 네트워크

어텐션 모듈은 합성곱 신경망에서 채널과 공간 정보에 가중치를 적용해, 향상된 특성을 추출을 제공하는 모듈이다. 채널 간의 상호 의존성을 모델링하여 중요한 채널을 강조하고, 미세먼지 농도 예측에 필요하지 않은 부분은 억제하는 역할을 해준다. 평균풀링(Average Pooling)과 최대풀링(Max Pooling) 연산을 모두 사용하여 입력받은 특징 맵의 정보를 수집한다. 두 종류의 Pooling을 사용하는 것은 컨텍스트 디스크립터(Context-Descriptor) F^c_{avg} 와 $F^c_{\max}$ 를 추출하여 서로 다른 공간의 특징을 수집하기 때문이다. 전역 평균/최대 풀링의 결과인 1D 벡터(Vector)를 입력으로 받는 FC-Layer를 사용한다. 이어서, 공유 네트워크(Shared-MLP)에서 채널 어텐션 맵($M_c \in R^{C\times 1\times 1}$)을 생성한다.

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (7) \\ &= \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{\max}))) \end{aligned}$$

즉, 각 디스크립터에서 공유 네트워크가 적용되면, 출력된 특징 벡터들을 요소별 합산(Element wise summation) 하여, 최종적으로 결합한다. 결합 된 특성 맵은 시그모이드(Sigmoid) 함수를 통해 결과를 (0, 1) 범위로 스케일링하여 각 채널의 중요도를 나타내는 가중치로 사용되며, 입력 데이터와 곱하여 받은 차원 크기와 똑같은 형태로 반환된다.

3.3 DNN 데이터를 통한 지역별 농도 예측

DNN 데이터는 대기질 측정자료 6개, 기상 측정자료 8개, CMAQ 예보자료 12개로 총 26개의 f특징으로 (?, 12, 26)처럼 구성된다. (?)는 총 날짜(=데이터 개수)를 뜻한다. 총 19개의 지역별로 DNN 데이터가 구

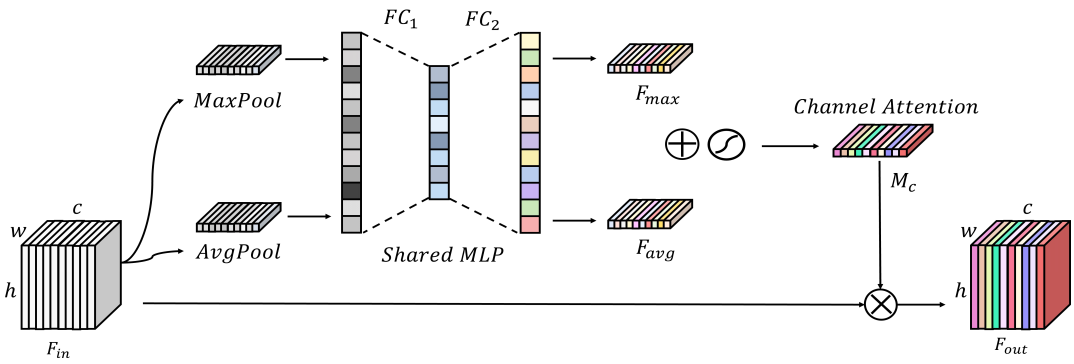

Fig. 3. Visualization for the process of channel attention.

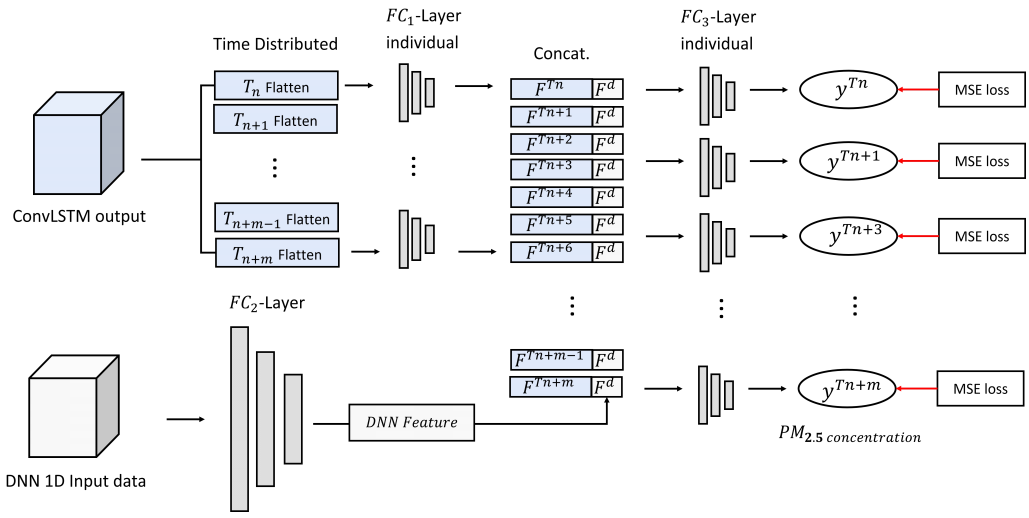

Fig. 4. Overall process of proposed ConvLSTM and DNN networks; note that DNN uses only one-dimension data for fine dust prediction.

성되어있으며, 모델에 입력하여 특정 지역에 대해 미세먼지 농도를 예측할 수 있다. 자세한 데이터 구성은 4.1에 설명되어있다. 입력 데이터는 ConvLSTM 네트워크를 통해 추출된 특징들과 결합(Concatenation) 되며 총 12개의 예측시간대로 분할 된다. 특징들은 각 예측시간별로 FC-Layer를 통해 (?, 12)의 예측농도를 산출한다. 제안하는 모델은 ConvLSTM 네트워크로 공간 이미지를 학습하고 DNN 네트워크와 정제된 데이터를 통해 지역별로 미세먼지 농도를 예측할 수 있다. 1차원적으로 미세먼지 농도의 패턴만을 학습하여 결과물을 산출하는 것이 아닌, ConvLSTM으로 공간 데이터를 학습하고 어텐션 모듈로 특징의 불필요한 부분을 억제하며, DNN 데이터로 지역별 학습 및 예측을 할 수 있어 관련 연구 분야 중 공간 데이터가 주로 사용되고 지역, 지점별 데이터가 보조로 사용되는 연구에서 활용될 수 있다고 생각된다.

4. 실험 결과 및 고찰

4.1 실험 환경 및 데이터셋

제안한 방법에 사용되는 데이터셋은 국가 및 지자체에서 운영하는 측정소에서 매시간별로 측정 기술(고층기상관측, 부유분진측정 등) 대기오염농도를 측정하여 국가대기환경정보 관리시스템(NAMIS)을 활용하여 수집한다. 본 연구는 국립환경과학원에서 제공하는 관측, 측정 기상 데이터를 사용하였다. 해당 데이터는 측정 시간대별로 분류 및 날짜 단위로

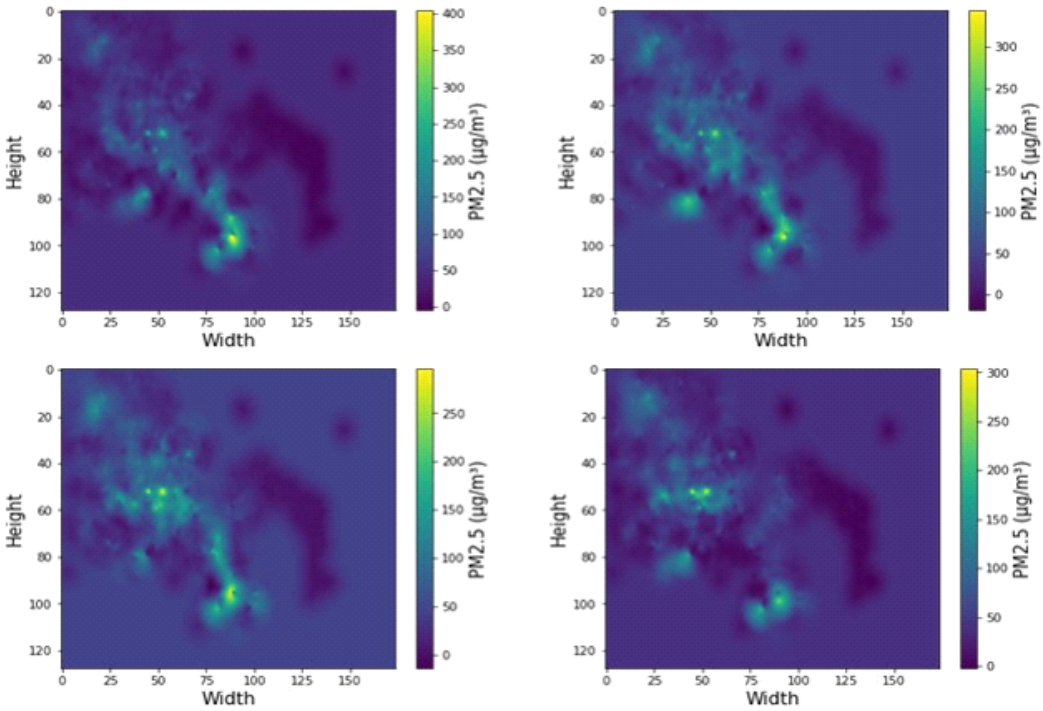

Fig. 5. Visualization example of $PM_{2.5}$ dataset, date by 2021-02-01~2021-02-04.

정리되어있으며 모델의 입력으로 사용되기 위해 전처리 과정이 필요하다. 전처리가 필요한 데이터는 PM_{10}, $PM_{2.5}$, O_3, NO_2, CO, 그리고 SO_2 등 농도가 포함되어 있으며, 시간별 기상 변수에는 풍속과 풍향, 상대습도, 기압, 대기 온도와 이슬점 온도, 복사량 및 누적 강수량이 포함되어 있다. 대기질 및 기상 관측소의 수는 연도별로 증가하고 있다. 2023년 기준, 대한민국에는 523개의 대기질 관측소와 103개의 기상 관측소가 있다. 제공되는 데이터가 아닌 개인 연구 개발 목적으로 공공 데이터를 사용하고자 한다면, .CSV 형식의 대한민국 대기질 데이터는 AirKorea 웹사이트(www.airkorea.or.kr)에서 얻을 수 있으며, 해당 데이터를 하루 6시간 단위로 미세먼지 농도를 순서대로 나열한다. 중국의 대기질 데이터는 PM2.5.in 웹사이트(pm25.in, 2022년 6월 이후 운영 중지)에서 얻을 수 있다.

관측된 대기질 및 기상 데이터는 크리깅 보간법(Kriging Interpolation)을 사용하여 주변 측정값들을 선형적으로 결합하여 값을 추정, 전처리를 진행하였다, Fig. 5에서 크리깅 결과 예시를 확인할 수 있다. 크리깅 보간법은 특정 관측 지점의 데이터를 결합하여 2차원 필드 데이터를 생성하는 지리 통계기법이다. 보간을 위해 주변 측정값들을 선형적으로 결합하고 이를 사용하여 값을 추정한다. 측정 지점 사이의 값을 추정하는 역거리 가중 보간법(Inverse Distance Weighting)[23]과는 달리, 주변에 있는 다양한 측정값들을 통계적으로 분석하여 값을 추정한다. 즉, 값을 추정할 때 크리깅 보간법은 측정값으로부터의 거리뿐만 아니라 각 이웃 측정값 사이의 상관관계 강도도 반영한다. 모델링 지역(중국과 대한민국을 포함한)의 관측 데이터를 크리깅 보간법을 이용해 이미지 데이터를 출력하고, 기존 (174, 128)의 이미지를 (90, 70)로 크기를 조정하여 2차원 그리드 데이터로 변환하여 사용한다(대한민국만 집중적으로 보기 위함).

학습 데이터는 2016년 1월 1일~2020년 12월 31일(7308개의 데이터를 사용), 검증 비율(Validation ratio)는 0.1, 테스트 데이터는 2021년 1월 1일~2021년 12월 31일(1460개의 데이터). DNN 데이터 또한 마찬가지로 학습 데이터는 2016년 1월 1일~2020년 12월 31일(7308개의 데이터를 사용), 테스트 데이터는 2021년 1월 1일~2021년 12월 31일(1460개의 데이터).

DNN 데이터는 19개의 지역에 대한 대기질 데이터 정보를 담고 있는 1차원 데이터로 학습 시 모델이 크리깅 된 데이터만 보고 예측하는 것이 아닌 실제 수치를 통해 예측하기 위해 입력되는 자료다. 데이터 셋에 포함되는 전체적인 데이터 정보는 Table 1에서 확인할 수 있다. 제안하는 모델은 예측 지역의 $PM_{2.5}$ 농도값을 예측한다. 예측 시간대는 03시, 09시, 15시, 21시로, 하루에 6시간 단위로 예측 시간 기준 현재, 내일, 모레의 D+00, D+01, D+02 농도가 예측된다. 입력과 출력의 길이는 (12), (12)로 진행하였지만, 유동적으로 변경할 수 있다. 미세먼지 예보모델은 텐서플로우(Tensorflow)와케라스(Keras)를 사용하여 구현되었다. 학습의 에폭 (Epoch)과 배치 사이즈(Batch size)는 각각 100과 10을 사용하였다. GPU는 RTX 2080 Ti를 사용하여 학습하였다. 본 논문에서는 서울 지역에 대해서만 평가 및 비교를 진행한다.

미세먼지 농도 예측을 위해 준비되어야 하는 1차원 DNN 데이터의 구성은 Fig. 6과 같다. $T_{+1} \sim T_{+15}$의 데이터는 학습 및 테스트 시, 수치모델 CMAQ과 성능 비교를 위한 자료로 입력된다.

미세먼지 농도 예측을 위한 19개의 지역에 대한 정보는 Table 2에서 확인할 수 있다. 총 5개의 광역으

Table 1. Configuration information of observation and forecast, air quality and meteorological variables used for input data.

type	Input Variables
Air quality observation	$SO_2(\mu m/m^3)$, $O_3(\mu m/m^3)$, $NO_2(\mu m/m^3)$, $CO(\mu m/m^3)$, $PM_{10}(\mu m/m^3)$, $PM_{2.5}(\mu m/m^3)$
Meteorological observation	$Ta(K)$, $Td(K)$, $Pa(hPa)$, $RH(\%)$, $U(m/s)$, $V(m/s)$, $RN_ACC(mm)$, $Radiation(0.01MJ/h-m^{2})$
CMAQ forecast	$T04 \sim T15f_PM_{2.5}$

SO_2(Sulfur dioxide), O_3(Ozone), CO(Carbon monoxide), NO_2(Nitrogen Dioxide), V and U(Horizontal Velocities), Td(Dew point temperature), Ta(Temperature), Pa(Pressure), Rad(Solar radiation), RN_ACC(Accumulative precipitation), $T04 \sim T15f_PM_{2.5}$(T04~T15 CMAQ Forecast), PM_{10} and $PM_{2.5}$(Particulate matter with aerodynamic diameters less than $10(\mu m/m^3)$ and $2.5(\mu m/m^3)$, respectively)

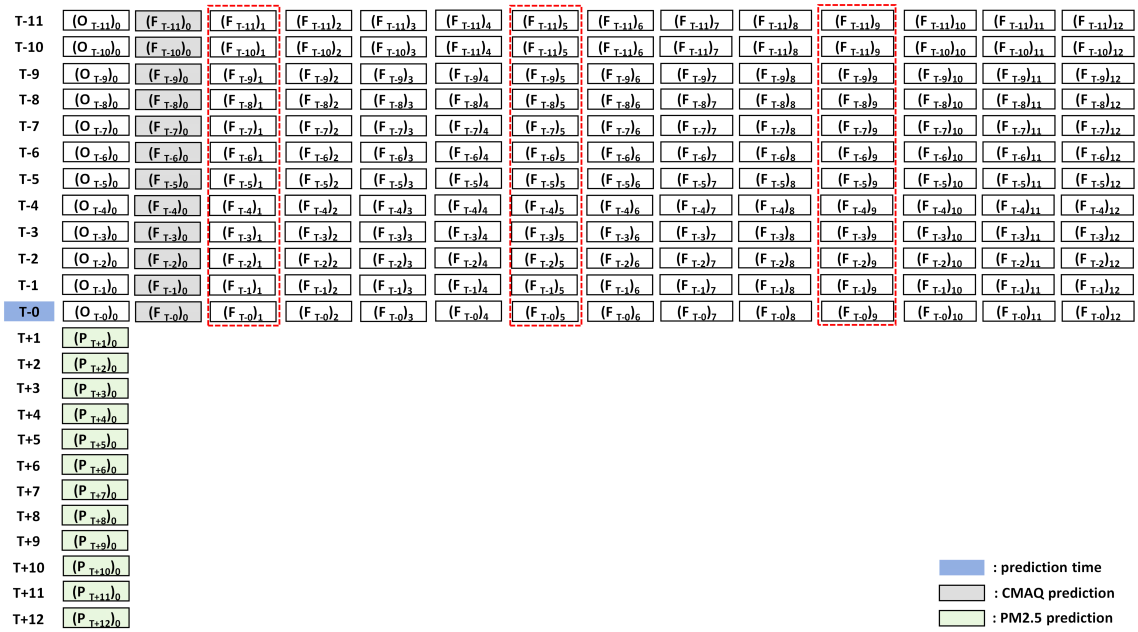

Fig. 6. DNN input data ($T_{-11} \sim T_{-0}$), ($T_{+1} \sim T_{+15}$) example.

Table 2. Regions and 19 areas for predicting fine dust concentration: SMA (Seoul Metropolitan Area).

Region	Area				
SMA+Gangwon West	Seoul	Incheon	Gyeonggi-Buk	Gyeonggi-Nam	Gangwon-YS
Gyeongsang-do	Daegu	Gyeong-Buk	Gyeong-Nam	Busan	Ulsan
Chungcheong	Daejeon	Sejong	Chung-Buk	Chung-Nam	
Jeolla-do	Gwangju	Jeon-Buk	Jeon-Nam		
Gangwon East	Gangwon-YD				
Jeju-do	Jeju				

로, 서울과 영서, 경상도, 충청도, 전라도, 강원 영동, 제주로 나누어져 있으며, 해당 광역에 속한 도시들을 지역이라 한다.

4.2 모델 평가

예보모델 결과와 측정자료의 정확도를 분석하는 방법으로 IOA(Index Of Agreement), MBIAS(Mean Bais), RMSE(Root Mean Square Error)를 사용하였으며, R(Correlation Coefficient) 항목을 추가하여 분석하였다.

$$IOA = 1 - \frac{\sum_{1}^{N}(Model - Obs)^2}{\sum_{1}^{N}(|Model - \overline{Obs}| - |Obs - \overline{Obs}|)^2} \quad (7)$$

$$R = \frac{\sum(Model - \overline{Model}) \times (Obs - \overline{Obs})}{\sqrt{\sum(Model - \overline{Model})^2 \times \sum(Obs - \overline{Obs})^2}} \quad (8)$$

$$MBIAS = \frac{1}{N}\sum_{1}^{N}(Model - Obs) \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{1}^{N}(Model - Obs)^2} \quad (10)$$

IOA는 일치 지수로, 예측값과 관측값 간의 일치도를 평가하기 위해 사용되는 통계적 지표로, 예측 모델의 성능을 측정하고 예측값과 실제 관측값 사이의 차이를 평가하는 데 사용된다. 수치가 1에 가까울수록 예측값과 관측값이 일치한다는 의미로, 반대로 0에 가까울수록 불일치의 정도가 크다고 설명할 수 있다. 상관 계수 R을 통해 예측값과 실제값의 선형 관계를 확인할 수 있다. 계수는 -1부터 1까지 범위를 가지며, -1은 완벽한 음의 선형 관계, 0은 관계가 없음을, 1은 완벽한 양의 선형 관계를 나타낸다. MBIAS는 평균 편향으로 예측값과 실제 값 간의 평균 차이를 나타낸다. 예측값과 실제 값 간의 차이를 합산하고 데이터 개수로 나누어 계산한다. 양(+)의 MBIAS는 모델이 실제 값을 과대평가하는 경향이 있음을 나타내며, 음(-)의 수치는 실제 값을 과소평가하는 경향을 나타낸다. 평균 제곱근 오차 RMSE는 예측값과 관측값 간의 차이를 나타내는 지표로, 예측 모델의 성능을 평가하는 데 사용된다. RMSE는 예측값과 실제값 간의 편차를 계산하여 예측 오차의 평균 크기를 측정합니다. 수치가 0에 가까울수록 모델의 성능이 우수하다고 할 수 있다.

$$ACC = \frac{A1 + B2 + C3 + D4}{N} * 100(\%) \quad (11)$$

$$POD = \frac{IV}{III + IV} * 100(\%) \quad (12)$$

$$FAR = \frac{II}{II + IV} * 100(\%) \quad (13)$$

$PM_{2.5}$에 대한 예보 성능은 지수에 대하여 적중률(ACC : Accuracy)을 분석, $PM_{2.5}$가 인체에 영향을 미치는 $35(\mu m/m^3)$를 초과하는 고농도 발생에 대해서는 감지 확률(POD : Probability of Detection), 거짓 경보율(FAR : False Alarm Rate)을 분석하였다. ACC, POD, FAR를 평가하기 위해서는 국립환경과학원에서 제공하는 공기 오염 수준에 따른 4가지 표현 단계가 필요하다. Table 3에서 나누어진 표현 단계별 그룹을 통해 관측(Observation)과 예측(Prediction)의 분포를 수식에 대입하여 예측성능을 산출한다.

4.3 실험 결과

최적의 어텐션 모듈 개수를 찾는 실험으로 어텐션 모듈의 개수를 1개(A), 2개(B)로 수정하여 최적의 어텐션 모듈 개수를 찾기 위해 실험을 진행하였다. ACC, POD, FAR로 결과를 추출하였다. 결과를 통해 알 수 있는 것은 (B) 네트워크가 단기적인 측면에서 더 나은 예측 성능 수치를 보여주고 있다. 오늘(D+00), T04 ~

Table 3. Air pollution level expression method for evaluating the forecasting performance of the forecasting model. This expression level was developed by National Institute of Environmental Research (NIER) in consideration of the anatomy impact and the sensible pollution class according to the air pollution level, which is expressed in 4 levels: good, moderate, bad, and very bad.

Type		Prediction			
		Good	Moderate	Bad	Very Bad
Observation	Good	A1	B1	C1	D1
	Moderate	A2	B2	C2	D2
	Bad	A3	B3	C3	D3
	Very Bad	A4	B4	C4	D4
I = [A1, A2, A3, A4], II = [B1, B2, B3, B4], III = [C1, C2, C3, C4], IV = [D1, D2, D3, D4]					

Table 4. The prediction performance when applying one (A) to two (B) attention module, bold means better.

	ACC		POD		FAR	
Time	A	B	A	B	A	B
T04	79.1	79.2	63.8	72.3	13.7	22.3
T05	73.5	76.5	53.7	68.1	16.5	21
T06	72.1	73.1	55.3	66.5	20	23.8
T07	70.6	70.5	62.2	68.6	27.3	30.6
T08	68.8	68.3	67.6	67.6	32.1	34.5
T09	67.5	68.4	61.2	64.9	35.8	34.1
T10	66.4	66.8	54	63.6	36.1	37.4
T11	65.3	66.2	56.1	55.1	41	36.8
T12	63	65.8	56.5	59.1	45	38.5
T13	61.4	65.2	61.3	53.2	49.1	38.5
T14	61.3	63.8	54.8	51.1	45.5	38.3
T15	56.3	60.2	52.2	40.9	51.5	43.3

T07 시간대에서 (B)가 좀 더 나은 ACC 성능을 보였으며, POD는 T11, T13, T14, T15를 제외한 전 시간대에서 같거나 조금 낮은 수치를 보였다. (A)보다는 낮은 수치를 보여주지만, FAR는 POD와 상충 관계(Trade-off)로서 (B)의 결과가 미세먼지 농도 예측을 (A)보다 더 정확하게 하며, 고농도 감지율 또한 높다고 볼 수 있다.

모델의 학습 및 검증 과정 중에도 (B) 모델이 손실률이 더 낮게 측정되고 있다. Fig. 8. 그래프를 통해 확인할 수 있다.

제안하는 모델과 비교하는 모델은 직접 구현한 모델들로 제공받은 데이터셋에 맞추어서 구현하였다. 여기서 CMAQ은 국립환경과학원에서 제공되는 프로그램을 사용하여 평가를 진행한다.

제안하는 모델이 어텐션을 적용하지 않은 Conv LSTM-DNN(w/o.) 모델과 비교하였을 때, 일치 지수 IOA가 0.1 증가, 상관 계수 R이 0.15 증가, 평균 편향 MBIAS가 1.27 감소, 평균 제곱근 오차 RMSE가 1.88 감소, 예측 정확도 ACC이 1.4% 증가, 탐지율 POD가 8.5%, 오경보율 FAR이 12.0%로 감소하였으며, 다른 모델과 비교했을 때, 미세먼지 예측성능이 높다고 평가할 수 있다. 미세먼지 농도 예측 연구에 가장 많이 사용되고 평가받아오는 LSTM 모델 또한 제안하는 모델이 평가 수치에서 더 나은 성능을 보여주었다. 제안한 모델과 수치 기반 예측 모델인 CMAQ과 비교하였을 때, 일치 지수 IOA가 0.13 증가, 상관 계수 R이 0.15 증가, 평균 편향 MBIAS가 1.82 감소, 평균 제곱근 오차 RMSE가 6.18 감소, 예측 정확도 ACC이 18.07% 증가, 탐지율 POD가 13.12%, 오경보율 FAR이 42.9%로 감소하여, 더 좋은 모델 성능과 예측을 보여준다. 수치 모델과 비교를 진행했을 때,



Fig. 7. T04~T15 fine dust prediction performance (ACC, POD, FAR) result with 1 and 2 attention module.

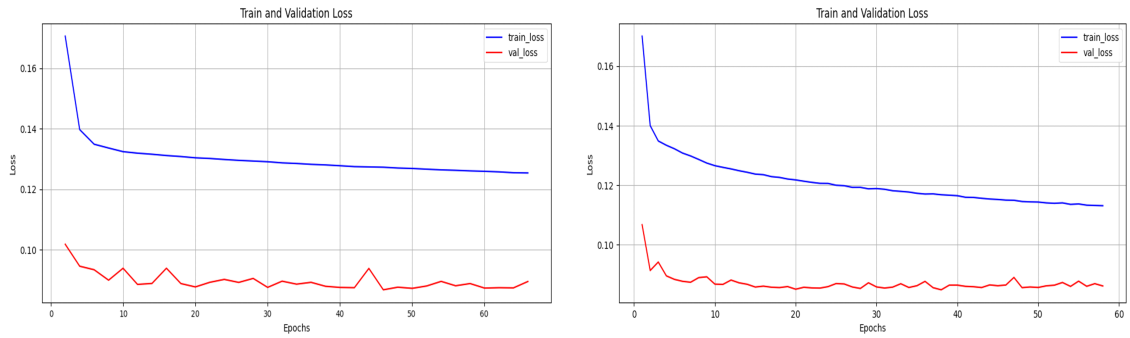

Fig. 8. Loss graph of attention models. (a) 'ConvLSMT-DNN(1 module)' and (b) 'ConvLSTM-DNN(2 module)'.

가장 중요한 미세먼지 예측 정확도 측면에서 약 18%가 상승하였음을 보여준다. 따라서 신경망 모델이 수치 모델보다 더 나은 예측성능을 보여준다고 할 수 있다. CNN-LSTM 모델 [24]은 CNN과 LSTM을 결합한 하이브리드(Hybrid) 모델로, 해당 논문에서 제시한 결과는 RMSE(17.9306)이었으며, 뛰어난 적중률을 보여주었다. ConvLSTM 네트워크와 가장 유사한 구조를 가진 네트워크로 비교 대상으로 가장 적합한 모델이라고 할 수 있다.

제안하는 모델은 CNN-LSTM과 비교했을 때, 일치 지수 IOA가 0.09 증가, 상관 계수 R이 0.12 증가, 평균 편향 MBIAS가 1.15 감소, 평균 제곱근 오차

Table. 5. Evaluation of proposed model, and comparison with other prediction model (attention with/without).

Model	IOA	R	MBIAS	RMSE
Proposed model	0.89/1	0.82/1	2.25	8.74
ConvLSTM-DNN(w/o.)	0.78/1	0.67/1	3.52	10.62
CNN-DNN(w.)	0.87/1	0.68/1	-2.57	9.23
CNN-DNN(w/o.)	0.85/1	0.66/1	2.67	9.98
CNN-LSTM[24]	0.80/1	0.70/1	3.40	11.63
WLSTM[4]	0.73/1	0.70/1	3.62	13.49
LSTM	0.75/1	0.72/1	3.01	12.95
CMAQ	0.76/1	0.67/1	4.07	14.92

Table 6. Prediction performance (ACC, POD, FAR) of proposed model, and comparison with other prediction model (attention with/without).

Model	ACC	POD	FAR
Proposed model	79.10%	82.40%	13.70%
ConvLSTM-DNN(w/o.)	77.70%	73.90%	25.70%
CNN-DNN(w.)	74.64%	80.39%	17.95%
CNN-DNN(w/o.)	73.15%	74.94%	19.37%
CNN-LSTM[24]	70.96%	75.28%	26.74%
WLSTM[4]	69.27%	65.62%	33.32%
LSTM	72.64%	72.15%	20.60%
CMAQ	61.03%	69.28%	56.60%

RMSE가 2.89 감소, 예측 정확도 ACC이 8.14% 증가, 탐지율 POD가 7.12%, 오경보율 FAR이 13.04%로 감소하였다. CNN-LSTM은 조금 부족한 성능 수치를 보여준다. 공간 이미지를 처리하고 특징을 추출하는 CNN과 시계열 패턴을 학습하는 LSTM 결합 모델이 미세먼지 농도를 예측하는 데 있어서 긍정적인 결과를 보여주지만, 시공간 데이터가 각각 CNN 레이어와 LSTM 레이어로 분리되어, 특징을 추출할 경우, 데이터가 손실되는 현상이(데이터가 복잡하여 상호 연관성 데이터를 놓치는) 발생한다. 따라서, 성능 저하를 야기 했다고 볼 수 있다. 하지만 ConvLSTM을 사용하여 각 레이어가 파라미터를 공유하고 상호작용을 통해 패턴을 인식하여 효율적으로 시공간 데이터를 동시에 처리할 수 있으며 제안하는 모델이 미세먼지 농도 예측에 최적화 되어 있음을 알 수 있다.

5. 결 론

본 논문에서는 공기 중 미세먼지 예측을 위해 주어진 공간 및 시간 데이터를 동시에 활용하는 ConvLSTM에 어텐션 메커니즘을 적용해, 전체적인 모델의 성능과 미세먼지 농도 예측 정확도를 향상시킨 딥러닝 모델을 제안하였다. 대기질, 기상 관측 및 측정 데이터를 신경망 모델에 입력 자료로 구성하여 ConvLSTM 구조에 입력으로 주어지는 공간 데이터를 어텐션 메커니즘에 활용하였다. 또한 시공간 정보를 모두 활용하여 모델이 기상의 변화 패턴을 학습하고, ConvLSTM 네트워크를 통해 추출된 특징과 DNN 지역별 1차원 데이터와 결합하여 오늘 D+00, 내일 D+01, 모레 D+02의 미세먼지 농도를 수치 모델(CMAQ)에 비해 18.0% 향상된, 80.0%에 근접하는 예측하는 정확도를 보여주었다. 수치 모델(CMAQ)[2]이나 미세먼지 예측에 사용되는 시공간 데이터 처리 CNN-LSTM 모델의 성능과 비교를 하였을 때도 향상된 예측 성능을 보여주었다. 실험 결과, 제안 모델은 기존 미세먼지 농도만 활용하여 예측하는 시계열 모델보다 높은 예측 정확도를 보였다. 이를 통해, 미세먼지 농도 예측을 위한 신경망 모델에 어텐션

메커니즘이 공간 및 시간적 데이터 예측 문제에서 유용하게 활용될 수 있음을 본 논문에서 보였다.

REFERENCE

[1] F. Lu, D. Xu, Y. Cheng, S. Dong, C. Guo, X. Jiang, et al., "Systematic Review and Meta-analysis of the Adverse Health Effects of Ambient PM2.5 and PM10 Pollution in the Chinese Population," *Environmental Research*, Vol. 136, pp. 196-204, 2015.

[2] US EPA Office of Research and Development. (2022), CMAQ (5.4), Zenodo, https://doi.org/10.5281/zenodo.7218076 (accessed June 28, 2023).

[3] Introduction to the CAI(2006), https://www.airkorea.or.kr/eng/khaiInfo?pMENU_NO=166 (accessed June 28, 2023).

[4] H. Hansika, H. Bergmeir, and K. Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions," *International Journal of Forecasting,* Vol. 37, Issue 1, pp. 388-427, 2021.

[5] F. Xiao, M. Yang, H. Fan, and A. Mohammed, "An Improved Deep Learning Model for Predicting Daily PM2.5 Concentration," *Scientific Reports*, Vol. 10, No. 1, 20988, 2020.

[6] S.K. Kim, J.M. Lee, J.W. Lee, and J.H. Seo, "Deep-dust: Predicting Concentrations of Fine dust in Seoul using LSTM," *arXiv Preprint*, arXiv:1901.10106, 2019.

[7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv Preprint*, arXiv:1409.1556, 2014.

[8] M. Ahmed, X. Zemin, and S. Yonglin, "Estimation of Ground PM2.5 Concentrations in Pakistan Using Convolutional Neural Network and Multi-Pollutant Satellite Images," *Remote Sensing*, Vol. 14, No. 7, 1735, 2022.

[9] X. SHI, Z. Chen, H. Wang, D.Y. Yeung, W. Wong, W. Woo, et al., "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," *Advances in Neural Information Processing Systems*, 28, 2015.

[10] S.H. Woo, J.C. Park, J.Y. Lee, and I. Kweon, "CBAM: Convolutional Block Attention Module," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3-19, 2018.

[11] J.M. Lim, "An Estimation Model of Fine dust Concentration using Meteorological Environment Data and Machine Learning," *Journal of Information Technology Services*, Vol. 18, No. 1, pp. 173-186, 2019.

[12] G.H. Lee and H.Y Kwon, "Improvement of PM2.5 Forecast by Categorical Wide Area Model," *Journal of Korea Multimedia Society*, Vol. 25, No. 3, pp. 468-475, 2022.

[13] B.Y. Kim, Y.K. Lim, and J.W. Cha, "Short-term Prediction of Particulate Matter (PM10 and PM2.5) in Seoul, South Korea using Tree-based Machine Learning Algorithms," *Atmospheric Pollution Research,* Vol. 13, Issue 10, p. 101547, 2022.

[14] E. Sharma, R.C. Deo, R. Prasad, A.V. Parisi and N. Raj, "Deep Air Quality Forecasts: Suspended Particulate Matter Modeling With Convolutional Neural and Long Short-Term Memory Networks," *IEEE Access*, Vol. 8, pp. 209503-209516, 2020.

[15] D. Shengdong, T.R. Li, and S. Horng, "Time Series Forecasting using Sequence-to-Sequence Deep Learning Framework," *2018 9th International Symposium on Parallel Architectures, Algorithms And Programming (PAAP), IEEE*, pp. 171-176, 2018.

[16] H.S. Hwang, J.Y. Lee, H.S. Seo, and S. Jeong, "A Study on the Effect of Fine Dust on Household Power Consumption using Climate Data - Focus on the Spring Season (April) and Fall Season (October) in Seoul," *Journal of the Society of Disaster Information*, Vol. 18, No. 3, pp. 532-541, 2022.

[17] J.Y. Byun, "Development of Particulate Matter

(PM2.5) Prediction Model for School Facilities Based on Living Lab Data," *Proceedings of the Korea Institute of Ecological Architecture and Environment*, pp. 90-91, 2022.

[18] K.Y. Nam, S.Y. Moon, and H.H. Kim "Analysis of the Correlation between Fine dust and Disease Using Big Data," *Proceedings of the Korea Information Processing Society Conference*, Vol. 29, No. 2 pp. 368-370, 2022.

[19] H. Yi, K.H.N. Bui, and C.N. Seon, "A Deep Learning LSTM Framework for urban Traffic Flow and Fine Dust Prediction," *Journal of Computing Science and Engineering*, Vol. 47 No. 3, pp. 292-297, 2020.

[20] F. Dempsey, "A Survey of Regional-Scale Blocking Patterns and Effects on Air Quality in Ontario, Canada," *Atmosphere*, Vol. 9, No. 6, 226, 2018.

[21] J.Z. Meng, "Raster Data Projection Transformation based-on Kriging Interpolation Approximate Grid Algorithm," *Alexandria Engineering Journal*, Vol. 60, Issue 2, pp. 2013-2019, 2021.

[22] Z.N. Yuan, Z. Xun, and T.B. Yang, "Hetero-Convlstm: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-temporal Data," *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* pp. 984-992, 2018.

[23] Z. Ren, L. Tian, and Y. Zhang, "Comparison of Kriging and Inverse Distance Weighting Methods for Soil Heavy Metal Interpolation in Chenzhou City, Hunan Province," *Chinese Journal of Geochemistry*, Vol. 252, pp. 459-467, 2011.

[24] T.Y. Li, H. Miao, and X.U. Wu, "A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5)," *IEEE* Access, Vol. 8, pp. 26933-26940, 2020.



이 준 민

2019년 북경제2외국어대학교 번역학과 학사
2021년~현재 한국외국어대학교 컴퓨터공학부 석사과정
관심분야 : 딥러닝, 패턴인식, 영상처리, 예측



김 경 태

2016년 중원대학교 학사
2017년~2019년 한국외국어대학교 컴퓨터공학부 석사
2019년~현재 한국외국어대학교 컴퓨터공학부 박사과정
관심분야 : 머신러닝, 패턴인식, 영상처리



최 재 영

2011년 KAIST 전기및전자공학과 박사
2008년~2012년 토론토대학 컴퓨터공학과 연구원
2012년~2013년 펜실베니아대학 의료공학과 연구원
2013년~2014년 삼성전자 책임연구원
2014년~2016년 중원대학교 의료공학과 조교수
2023년~2024년 브리티시컬럼비아대학 컴퓨터공학과 방문교수
2016년~현재 한국외국어대학교 컴퓨터공학부 부교수
관심분야 : 딥러닝, 머신러닝, 패턴인식, 영상처리