Short-term PM2.5 Prediction using Modified Attention Seq2Seq BiLSTM

Ida Bagus Krishna Yoga Utama, Duc Hoang Tran, Yeong Min Jang
Department of Electronics Engineering
Kookmin University
Seoul, South Korea
Email: idabaguskrishnayogautama@gmail.com; duchoangbkdn.1995@gmail.com; yjang@kookmin.ac.kr

***Abstract*—In semiconductor industry, concentration of particulate matter in cleanroom is important to maintain as it can damage the wafer die in the manufacturing process. The environment in semiconductor industry is well maintained where the temperature and humidity are always stable. Hence, it is not possible to use other features to predict the PM2.5 concentrations. In this paper, we present a modified attention Seq2Seq BiLSTM model for predicting the PM2.5 concentration. Based on the proposed model, short term (60 minute ahead, 90 minute ahead, and 120 minute ahead) PM2.5 concentration predictions was built. The proposed model also compared with Seq2Seq LSTM, Seq2Seq BiLSTM, and attention Seq2Seq LSTM models where the proposed model outperformed those models by achieving lowest RMSE, MAE, and MAPE values in all prediction time length.**

Keywords—Attention Seq2Seq BiLSTM, particulate matter, PM2.5, IoT platform, prediction

I. Introduction

It's been known for a long time that air pollution is harmful for human health. The air pollution is caused by particulate matter (PM) that comes from varieties of sources. PM is mixed of microscopic solid particles, such as dust, dirt, and liquid droplets poses more danger to human health [1]. Generally, PM is separated into three main groups: Coarse PM, Fine PM, and Ultrafine PM. Coarse PM is PM that having a diameter of 10 um and smaller that can irritate eyes, nose, and throat. Then, Fine PM is PM that having a maximum diameter of 2.5 um and can enter deep into the lungs and even the bloodstream. Lastly, Ultrafine PM is PM that having a diameter of 0.1 um and smaller, it is considered ad the most dangerous PM because it can infiltrate our bodies to an even greater extent [2]. Due to the harmful effects of air pollution, many researchers conduct studies on prediction of air pollution. However, real-time air pollution prediction is difficult due to the time-series data which are usually, incomplete, complex, and non-linear [3].

The air pollution prediction research is mostly conducted for urban area applications where the objectives is predicting the future air pollution concentration in a city. Tiwari et al [4] propose a LSTM model for predicting air quality in Delhi. Xayasouk et al [5] employs LSTM and deep autoencoder models to predict South Korea's air pollution where the prediction focused on PM10 and PM2.5 concentrations. Zhang et al [6] shows promising result on predicting the air quality in urban area. Their prediction model employs multivariate features which includes PM concentration, NO2, CO, O3, traffic data, and meteorology data. Huang [7] proposes a deep neural network model called APNet which based on CNN-LSTM for prediction PM2.5 concentrations. Their prediction result proves that the proposed APNet can achieve the best performance compared to other machine learning methods. In their study, they train the model by using combination data of PM concentration, wind speed, and hours of rain.

Air pollution prediction is also applied in industrial field, especially semiconductor industry. In semiconductor applications, air pollution prediction is used to estimate the concentration of particulate matter. Estimating the concentration of particulate matter is important because the wafer die produced in the semiconductor industry should be free from any contamination, otherwise the wafer die will be unusable. Multiple sources could generate organic or inorganic particulate matter, such as scratches, fractures, overlay flaws, and stress where those process could happen in industrial environment, tools, or processes [8].

Aji et al [9] presents a multi-dense layer BiLSTM model for predicting the PM2.5 concentration in semiconductor industry environment. In their work, they predict 1 hour ahead, 2 hour ahead, and 3 hour ahead and yield a good prediction result proven with low error value compared to other algorithms. Unlike predicting PM2.5 concentration in urban area, in semiconductor industry, the environment is being kept at a constant state. Therefore, the environment value like temperature and humidity in the room is always stable. As such, predicting the particulate matter in semiconductor industry is only feasible by using only particulate matter data because we cannot capture data of scratches, fractures, overlay flaws, and stress that occurred.

In this work, we present a technique of predicting PM2.5 concentration by using modified sequence-to-sequence BiLSTM with attention method. The proposed model is used for predicting 60 minute ahead, 90 minute ahead, and 120 minute ahead of PM2.5 concentrations. The approach used in this work is multiple output strategy where the entire prediction sequence is predicted in one-shot manner.

II. System Overview

A. Data Collection

To provide PM2.5 data for deep learning model training, we develop IoT platform to collect PM2.5 data from the environment. The overall architecture of the developed IoT platform is described in Figure 1. To sense the PM2.5 concentration, DOTECH PSU650 sensors is used in this work. The PSU650 sensor is connected to a PLC by using Modbus RTU connection. The sensor send data to the PLC every 1 second. Then, the PLC will relay the PM2.5 data to the edge server via MQTT. In this work, NVIDIA Jetson Nano is used as edge server that has two functions: storing the PM2.5 data into local database and execute real-time PM2.5 predictions. The local database is using MongoDB to store PM2.5 data every 1 second.

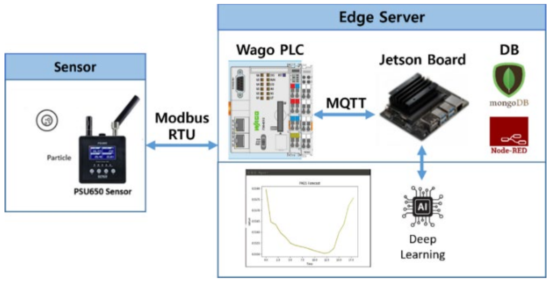

Figure 1. Developed IoT Platform for PM2.5 data collection

III. Proposed Framework

To predict the PM2.5 concentrations, we propose a modified sequence-to-sequence BiLSTM with attention model that only using PM2.5 data as input features. We use 15 days of PM2.5 data collected by the developed IoT platform with data points date from 10 March 2022 to 24 March 2022. Although we collect data every 1 second, for developing the model we resample into every 1 minute data due to highly oscillating value in 1 second data. As such, we will use a total data of 21,600 data points for developing the model.

A. *Data Preprocessing*

The collected PM2.5 data is highly fluctuating, to make the model learn better, we need to do several preprocessing steps. First, we analyze existence of null value in the data. Null value can occur when some errors happen in the IoT platform. To resolve null value issue, we use the previous value to fill the null value.

$$Y_{diff} = Y_t - Y_{t-1} \tag{1}$$

Then, we applied de-trending method to the data by subtracting the current data points with previous data points with equation denotes by Equation 1. The objective is to remove trend in the data, and it supposed to enable the model learns about the data easier. Next, we apply normalization towards the de-trended data and rescale the data into -1 to 1 value, to fit with the Tanh activation function that will be used in the model. Equation 2 shows the formula to normalize the data.

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

In this work, we only use PM2.5 data to make a prediction. However, using only single feature cannot produce decent results due to lack of information in the data. Hence, we add additional features in the data by using lagged value from the data. We use 300 lagged features as additional features, and the data has total features of 301 features. To create sequence-to-sequence, we convert the data into sliding window format. The sliding window data will consist of the past 300 minute data points and future 60 minute, 90 minute, and 120 minute data points. Finally, we separate the data into three parts, train data, validation data, and test data, with size of 70%, 20%, and 10%, respectively.

B. *Modified Sequence-to-Sequence BiLSTM with Attention*

The sequence-to-sequence model (Seq2Seq) originally proposed by Sutskever et al [10] to tackle the task of translating a sequence into another. Over the years, Seq2Seq has been widely used in processing task of variable input and output sequences, including machine translation, speech recognition, and time-series prediction. Seq2Seq use encoder decoder architecture where it can map a variable input sequence into a variable length output sequence.

Seq2Seq has two main parts, it is encoder and decoder. The encoder will receive input in forms of a sequence and from each input sequence $x(x_1, x_2, \dots, x_t)$, it will change its hidden state accordingly. The encoder part will output a corresponding hidden state $h_{enc\ t}$ which each hidden state time step $h_{enc\ 1}, h_{enc\ 2}, \dots, h_{enc\ t}$ will be a vector C. For each time step, vector C, the previous decoder output $Y_{t\prime}$, and the hidden state $h_{dec\ t\prime}$ is used as input for the decoder. The decoder receives those vector and process the similar way as in encoder to output an output sequence. The encoded sequence from encoder contains all information of the original sequence. Sometimes the amount of information is too large which can deteriorate the performance of the model. Hence, attention mechanism is introduced to enable the model to focus only to certain part of the vector C. The attention vector $a_{t\prime t}$ is calculated by using equation 5

$$h_{enc\ t} = f(x_t, h_{enc\ t-1}) \tag{1}$$

$$C = f(h_{enc\ 1}, h_{enc\ 2}, \dots, h_{enc\ t}) \tag{2}$$

$$Y_{t'} = f(Y_{t'}, C) \tag{3}$$

$$h_{dec\ t\prime} = f(Y_{t'-1}, h_{dec\ t\prime-1}) \tag{4}$$

$$a_{t\prime t} = \exp\left(score(h_{dec\ t'}, h_{enc\ t})\right) \cdot \sum_{k=1}^{T} \exp\left(score(h_{dec\ t'}, h_{enc\ t})\right)^{-1} \tag{5}$$

In this work, we employ BiLSTM in the encoder and LSTM in the decoder of the modified Seq2Seq model. The modified Seq2Seq model consist of one encoder and two decoder. By using BiLSTM, we can get two types of hidden state data, forward state, and backward state. The forward state will be used as input for the forward decoder. Meanwhile the backward state will be used for the backward decoder. Due to using two decoder, we will have two outputs. Then, we concatenate the two outputs from the forward decoder and backward decoder and apply dense layer to calculate the combined output. The dense layer will have the same number of units as the length of the output sequence. Figure 2 describe the complete architecture of the modified attention Seq2Seq BiLSTM model.

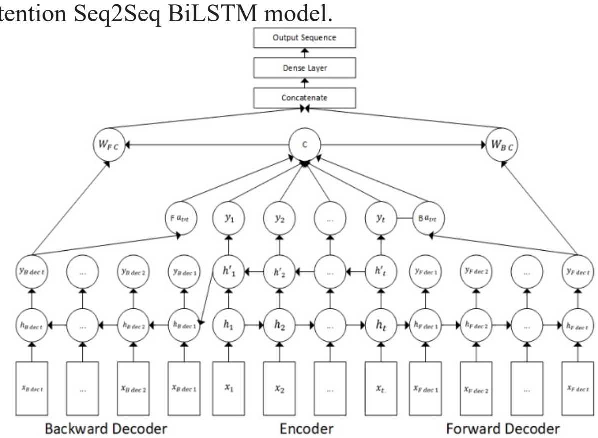

Figure 2. Architecture of the modified attention Seq2Seq BiLSTM model

IV. RESULTS AND DISCUSSION

After the data is ready and the model has been designed, the next step is we train the proposed model using the prepared data. The environment for training the model is using hardware and software as described in Table 1.

Table 1. Environment for training the proposed model

Processor	Intel Xeon Silver
GPU	RTX 3090
RAM	128GB
Library	PyTorch 1.11.0
Language	Python

In the training process, MSE loss is used as the criterion function. To optimize the training performance AdamW optimizer is applied. Then, the learning rate is using value of 0.001 and total epoch is 100. For this work, we compare the performance of the proposed method with three other model: attention Seq2Seq LSTM, Seq2Seq LSTM, and Seq2Seq BiLSTM. All model is being trained in the same environment with same number of epoch but different hyperparameter. All models will be used to do PM2.5 prediction for 60 minute ahead, 90 minute ahead, and 120 minute ahead.

Table 2. Training results

Model	Prediction Time Length	RMSE	MAE	MAPE
Attention Seq2Seq LSTM	60 minute	**66.639**	88.622	0.267
	90 minute	**96.03**	80.483	0.397
	120 minute	221.082	189.975	0.81
Seq2Seq LSTM	60 minute	148.249	111.675	0.28
	90 minute	137.771	109.544	0.426
	120 minute	216.392	178.258	0.743
Seq2Seq BiLSTM	60 minute	172.357	127.172	0.412
	90 minute	114.478	96.095	0.428
	120 minute	139.078	106.204	0.429
Proposed model	60 minute	91.8266	**70.052**	**0.263**
	90 minute	98.227	**71.067**	**0.304**
	120 minute	**95.592**	**72.298**	**0.294**

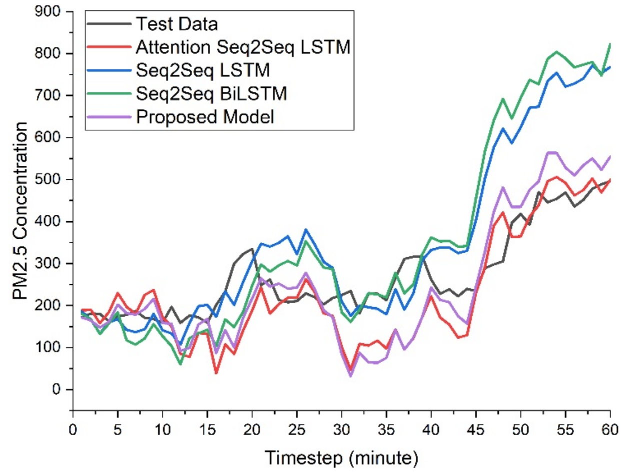

Figure 3. 60 minute ahead prediction result

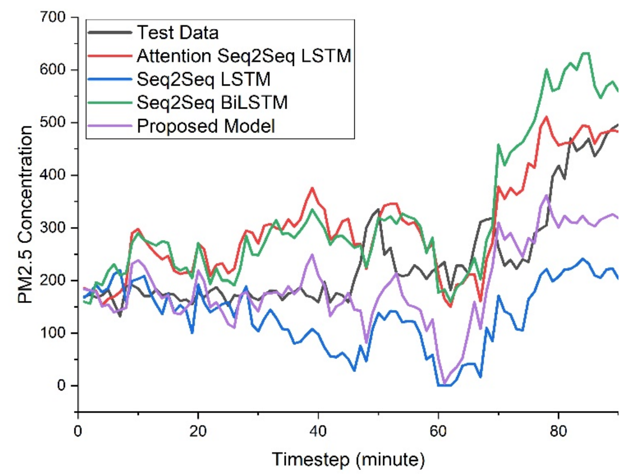

Figure 4. 90 minute ahead prediction result

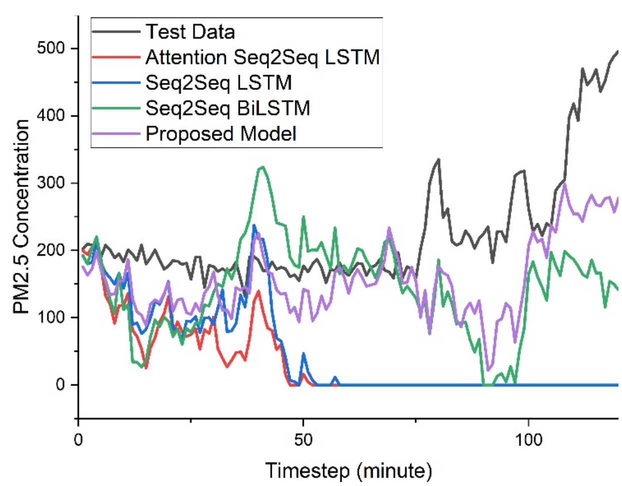

Figure 5. 120 minute ahead prediction result

From table 2, it shows the error measurement of all models for three prediction time length. From the table, we can see that in 60 minute ahead prediction, all four models can produce similar value where the RMSE, MAE, and MAPE produced by all models doesn't have much difference. This means that for 60 minute ahead prediction, all four models still can predict the PM2.5 concentration accurately. However, the proposed modified attention Seq2Seq BiLSTM still leads with lowest MAE and MAPE measurement error. Then, for 90 minute ahead prediction, the measurement error difference between four models is starting to have larger differences. The Seq2Seq LSTM model produce worst measurement error for 90 minute ahead prediction with high RMSE and MAE compared to other algorithms. The Seq2Seq BiLSTM is slightly better than Seq2Seq LSTM, and the attention Seq2Seq LSTM is better than Seq2Seq BiLSTM. But model that achieve best measurement error in 90 minute ahead prediction still the proposed model. For the 120 minute ahead prediction, the advantages of having two decoder in the proposed model is seen. The proposed model still able to produce prediction with value similar to the ground truth data. Meanwhile other algorithm, attention Seq2Seq LSTM and Seq2Seq LSTM cannot produce decent prediction by only generating 0 value after several timesteps. We can also see that using attention is advantageous to the model which the attention Seq2Seq LSTM and the proposed model can produce better prediction result than other model that doesn't utilize attention. Also, using bidirectional in BiLSTM is improving the model performance compared to only single direction.

From those results, we can conclude that the proposed model is superior to other three model, especially for longer time length prediction. This performance can be achieved due to the help of using two separated decoder and separated attention. The separated decoder can help the model to retain specific information for forward sequence and backward sequence. In normal BiLSTM, usually the forward sequence is neglected although it still contains some information. In the proposed method, all information is retained and used for the prediction. Also, the separated attention for each encoder helps the model to keep focus on certain important part of the sequence.

V. Conclusion

This paper proposed a modified version of attention Seq2Seq BiLSTM for predicting PM2.5 concentration in semiconductor industry where the prediction only utilize PM2.5 data. The proposed model is used for predicting 60 minute ahead, 90 minute ahead, and 120 minute ahead of PM2.5 concentration. Due to the application of two decoder for forward sequence and backward sequence and application of attention mechanism, the model can achieve best performance compared to other model, especially for prediction with longer time length.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2022-2018-0-01396) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

References

[1] Kim, K.-H., Kabir, E., & Kabir, S. (2015). A review on the human health impact of airborne particulate matter. Environment International, 74, 136–143. https://doi.org/10.1016/j.envint.2014.10.005.

[2] Smith, D. (n.d.). The three types of particulate matter: All about PM10, PM2.5, and PM0.1. Kaiterra.Com. Retrieved May 14, 2022, from https://learn.kaiterra.com/en/resources/three-types-of-particulate-matter.

[3] Oprea, M., Popescu, M., Dragomir, E. G., & Mihalache, S. F. (2017). Models of particulate matter concentration forecasting based on artificial neural networks. 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS).

[4] Tiwari, A., Gupta, R., & Chandra, R. (2021). Delhi air quality prediction using LSTM deep learning models with a focus on COVID-19 lockdown. In arXiv [cs.LG]. http://arxiv.org/abs/2102.10551.

[5] Xayasouk, T., Lee, H., & Lee, G. (2020). Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. Sustainability, 12(6), 2570. https://doi.org/10.3390/su12062570.

[6] Han, Y., Zhang, Q., Li, V. O. K., & Lam, J. C. K. (2021). Deep-AIR: A hybrid CNN-LSTM framework for air quality modeling in metropolitan cities. In arXiv [cs.LG]. http://arxiv.org/abs/2103.14587.

[7] Huang, C.-J., & Kuo, P.-H. (2018). A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors (Basel, Switzerland), 18(7), 2220. https://doi.org/10.3390/s18072220.

[8] The International Technology Roadmap for Semiconductors 2.0. Available online: https://www.semiconductors.org/wp-content/uploads/2018/06/4_2015-ITRS-2.0-ESH.pdf (accessed on 30 December 2021).

[9] Prihatno, A. T., Utama, I. B. K. Y., & Jang, Y. M. (2022). OneM2M-enabled prediction of high particulate matter data based on Multi-Dense Layer BiLSTM model. Applied Sciences (Basel, Switzerland), 12(4), 2260. https://doi.org/10.3390/app12042260.

[10] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with Neural Networks. In arXiv [cs.CL]. http://arxiv.org/abs/1409.3215