

An Air Pollution Prediction Scheme Using Long Short Term Memory Neural Network Model

Jongchol Kim^{1,*}

¹School of Humanities & Law, Northeastern University, Shenyang 110169, China School of Control Science, University of Science, Pyongyang 999093, Democratic People's Republic of Korea

Abstract—In order to establish countermeasures for air pollution, it is first necessary to accurately grasp the air pollution state and predict the cause and change trend of the pollution situation. Due to the continuously strengthening regulations on the emissions of environmental pollutants, the forecasting and management of nitrogen oxides (NO_x) emissions is receiving a lot of attention from industrial sites. In this study, a model for predicting nitrogen oxide emissions based on artificial intelligence was proposed. The proposed model includes everything from data preprocessing to learning and evaluation of the model, and used a Long Short-Term Memory (LSTM) neural network model, one of the recurrent neural networks, to predict NO_x emissions with time-series characteristics. The optimized LSTM model showed more than 93% NO_x emissions prediction accuracy for both the training data and the evaluation data. The model proposed in this study is expected to be applied to the development of a model for predicting the emission of various air pollutants with time-series characteristics.

1 Introduction

With the development of the modern industry, which uses the fossil fuels as basic energy, an air pollution caused by air pollutants emitted during combustion of fossil fuels has worsened. In order to detect the degree of air pollution and establish improvement measures, it is important to detect the amount of air pollution and predict the degree of pollution in the future. Currently, the emissions of the air pollutants are collected in many countries, and standard pollutants such as SO₂, NO₂, O₃, PM₁₀, PM_{2.5}, and CO have been measured at air measurement sites. In particular, nitrogen oxide (NO_x) is a term collectively referred to as nitrogen oxides such as nitrogen monoxide (NO), nitrogen dioxide (NO₂) and dinitrogen pentoxide (N₂O₅), and is mainly generated in the process of burning fuel in automobiles or industrial sites. When NO_x is released into the air, it generates fine dust and ozone through the photochemical reaction of sunlight, so it has been designated as the subject of environmental regulations internationally [1,2]. As the air pollution problem is important, studies have been conducted to forecast the degree of pollution through various research methods.

Traditional time series forecasting models include Auto-Regressive (AR), Moving Average (MA), and Auto-Regressive Integrated Moving Average (ARIMA). In particular, the

ARIMA model outperformed other time series models such as AR and MA in the prediction accuracy [3,4]. Kohzade et al. compared the ARIMA model with an artificial neural network (ANN), and Ho and Aladag et al.

compared the performance of the ARIMA model and Recurrent Neural Network (RNN) [5,6]. However, since these models assume the relationship between time series data as a linear relationship, there is a disadvantage of poor performance when applied to real data having nonlinearity [7,8]. McKendry and Zhao compared the multilayer perceptron and the regression algorithms, and reported that the single regression algorithms showed better performance, but did not show good predictive performance [9,10]. As deep learning has recently attracted attention, various networks and algorithms are being utilized and developed. Among them, LSTM neural network proposed by Hochreiter et al. was developed for the analysis of text or speech as a kind of RNN [11]. The main characteristic of the LSTM neural network is that it helps predict future values by properly preserving data from the past in the LSTM neural network through input, forget and output gates. Due to this structural advantage, this method is a very widely used deep learning method for predicting of the time series data. RNN models have high predictive performance for nonlinear data and provide competitive results compared to traditional time series prediction models [12-14]. In this work, in order to consider the time dependence of NO_x emissions, a model for predicting of the NO_x emissions was developed using a recurrent neural network. The proposed model covers from data preprocessing to RNN model training, optimal parameters and evaluation. The paper is presented as follows. In Section 2, the circulatory neural network and LSTM model used to handle time series data are described. In Section 3, the data used for verification were introduced, and prediction performance was compared by selecting

*e-mail: jinzhengzhe922@163.com

neural network parameters and comparing predictions through empirical data. Finally, the conclusions of this paper are summarized and further studies are discussed.

2 Deep Learning Algorithm

2.1 The Recurrent Neural Network

A recurrent neural network is a kind of artificial neural network, and is a neural network used to manipulate time series data. Figure 1 shows the structure of the recurrent neural network. When the structure on the left is unfolded according to the passage of time, it is unfolded in the shape of the right side of the arrow. In figure 1, x represents the input layer, O represents the output layer, and h represents the hidden layer. And W , U , and v represent the weight of the connection, respectively. Looking at the process of the recurrent neural network, the input enters the hidden layer through the input layer over time.

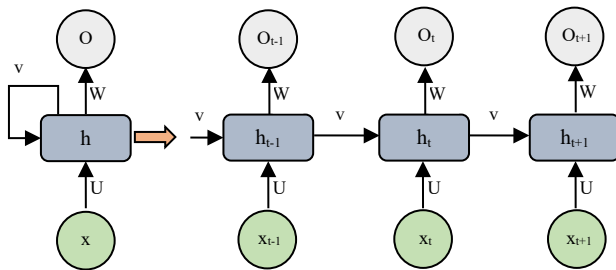


Figure 1 Structure of recurrent neural network

The output value of the previous hidden layer goes into the input of the current hidden layer, and the output value is output based on these two values. This process is repeated until all inputs are entered.

2.2 LSTM Neural Network

The LSTM network is a special kind of recurrent neural network, and is a network that uses LSTM blocks as a

hidden layer. The LSTM block is a structure in which the state value of the previous block enters the current block and is involved in the operation, like the structure of the recurrent neural network. A typical LSTM block is configured mainly by three gates: forget gate, input gate, and output gate. Figure 2 shows the structure of an LSTM block. Now with Figure 2, the role of each gate will be described in order [15]. The first step is the forgetting door. The hidden layer outputs of the current and previous inputs go through this layer, denoted by σ . Choose the amount of information that will go through this layer. σ represents the sigmoid activity function. The mathematical expression of this function is as follows [16-18].

$$S(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

The value obtained through this function is between 0 and 1. In the output value, 1 means to pass all the input values of the function, and 0 in the output value means to delete all input values. The output of the forget gate is as follows.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (2)$$

Here, W denotes the weight of the layer, and b denotes the bias value of the layer. The second is the input gate. This is the process of selecting what value to add to the value passed through the forget gate. The candidates for the newly added value are determined by the output of the previous hidden layer and the output of the tanh activation function layer of the current input. The output of the tanh function has a value between -1 and 1. The candidate of the new value and the value of the cell state are as follows.

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\bar{c}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (4)$$

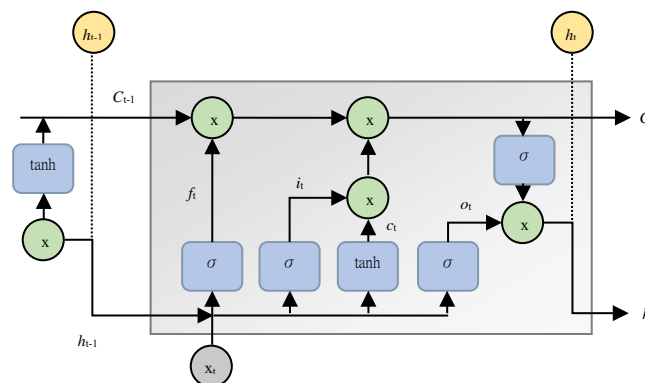


Figure 2 Structure of LSTM block

Based on these two values, we need to update the state value of the current block. The operation can be expressed as an equation as follows.

$$C_t = f_t * C_{t-1} + i_t * \bar{c}_t \quad (5)$$

Finally, the output gate determines the output of this block. First, the sigmoid function determines the state of the new output. Then, the block state value calculated in the previous step is passed through the tanh layer to make a new output candidate again. The product of these two values exits the block as the output of the next hidden layer and the current hidden layer. This can be expressed as an equation as follows.

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

Through this process, the LSTM block receives input, processes the block status value, and sends out an output.

2.3 Supervised Learning

Methods of machine learning include supervised learning, unsupervised learning, and reinforcement learning. Since supervised learning is used in this paper, supervised learning is briefly described. Supervised learning is a learning method that informs the input data and the correct answer to determine the learning direction when training a model. The model only accepts inputs and compares the corresponding outputs. Based on this error value, the bias value and weight in the model are adjusted. Schematic of this process is shown in figure 3.

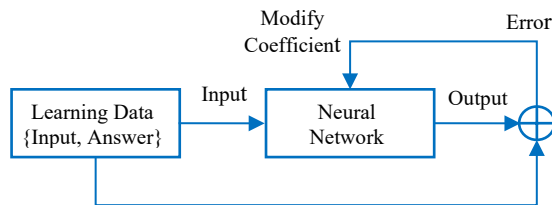


Figure 3 Structure of supervised learning.

At this time, the method of expressing the error value varies depending on the loss function used. And based on this loss function, the optimization algorithm is the algorithm that allows the various coefficients in the model to quickly find the optimal value. There are various loss functions and optimization algorithms in machine learning. The learning speed and performance of the model are different depending on which function and algorithm are selected. The loss function and optimization algorithm of the model used in this paper are Mean Squared Error (MSE) function and Adam optimizer, respectively [19].

3 EMPIRICAL RESULT AND DISCUSSION

3.1 Experimental Setup

In the work, an experiment was conducted on a prediction method based on and deep learning through empirical data. As empirical data, Air Pollution: Real-time Air Quality Index (AQI) from the 2005 to 2019, yearbooks provided

at <http://aqicn.org/city/london/> was used. The data used are shown in figure 4.

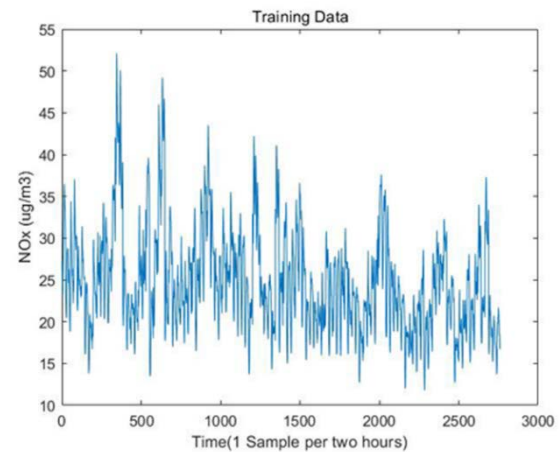


Figure 4 Empirical data for training.

When the distribution difference between the data is large, the MinMax Scalar was used for all data except the predicted value to prevent the load from being trained. The range of the data was matched with a value between 0 and 1. In order to obtain a good predictive model, it is important to set the values of the parameters of the LSTM network. In LSTM network, there are many hyperparameters such as learning rate, optimization method, dropout rate, epoch, and number of hidden. If the learning rate is low, the speed of learning is slow, and if the learning rate is large, learning may not be performed and vibration may occur. It is usually specified as a value between 0.0001 and 0.1. In this paper, it was set as 0.01. As an optimization method, the ADAM method was used, and the dropout ratio was set as 0.5.

In this paper, the epoch and the number of hidden units are not arbitrarily determined, but the epoch and hidden units with the lowest MSE are used through several combinations. The method is as follows.

- a) Divide the data into training, validation, and test data.
- b) The seed is fixed, and the model is trained with the training data for a all combinations of the number of hidden units and epoch.
- c) Verify the learned model using the verification data. At this time, check the loss change according to the epoch and check whether overfitting or underfitting occurs.
- d) Select the number of hidden units with the lowest MSE and the epoch.
- e) Create new training data by combining the training data and verification data in step (1).
- f) The model is trained using the selected number of hidden units and the epoch, and predictions are made based on this.

In the above process, the test data are not used to train or verify the model. In other words, the test data is only used when comparing it with the predicted value.

The training of LSTM was implemented on CPU: Intel Core i5-8400H 2.5GHz, and GPU: GTX1660 with

memory 12G. Software tools used include MATLAB R2019a and python.

3.2 Experimental Result

Based on the training data, various learning parameters and model structures were tested. For the evaluation of the model, the Root mean square error (RMSE) of Eq. (8) was used and is as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (8)$$

Here, y_i is the actual value of the i -th sample, \hat{y}_i is the predicted value of the i -th sample, and n is the number of samples. As mentioned above, the learning parameters adjusted through the experiment are the number of LSTM layers, the number of LSTM cells, epoch, and batch size. First, it is the number of LSTM layers. In this work, experiments were conducted on one-, two-, and three-layer structures. One and two layer structures showed performance that could not keep up with detailed data, but three layer result was satisfactory. Next is the number of LSTM cells. This parameter determines how many LSTM blocks are set in the LSTM hidden layer.

In the work, the optimal value was set to 64 through an experiment. Even if the same data is learned, the performance of learning tends to increase when iterative learning is performed. However, raising this value can lead to overfitting. So, in this study, 100 epochs were set. The batch size is a parameter that determines how many datasets to be trained. If this value is too large, it was judged that it would be difficult to follow detailed data, and this value was set as 32 through an experiment. The training was conducted through the set data and parameters, and the output of the trained model was checked by inputting the test data. The results are shown in figure 5.

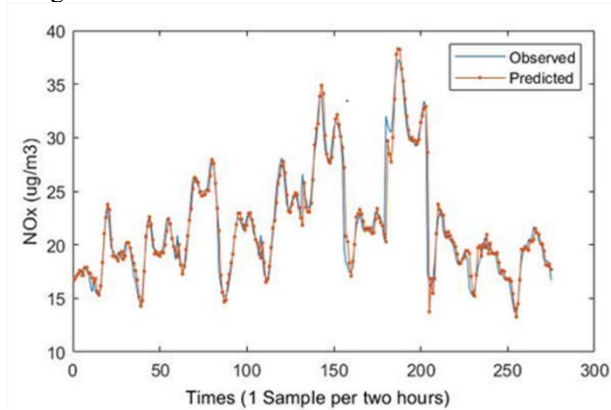


Figure 5 Prediction result and test data.

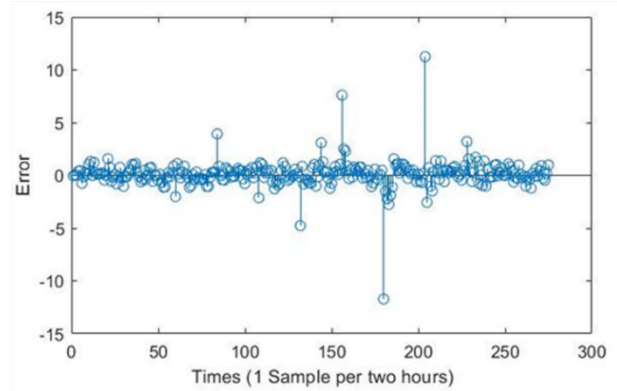


Figure 6 Error of prediction result

The accuracy of the NOx prediction model was confirmed through the plot in figure 6, and figure 6 shows that the learning model with an accuracy of 0.93 was secured as a result of learning with the training data. From these results, it was confirmed that a NOx prediction model with high accuracy was developed for not only the training data but also the evaluation data.

4 CONCLUSION

In this work, the concepts of deep learning, circulatory neural networks, and LSTM in artificial neural networks and supervised learning as a learning method were investigated. In addition, a model for predicting NOx emissions was presented using LSTM, an artificial intelligence technique, and its validity was verified by applying it to empirical data. The LSTM based NOx prediction model was trained using the training data, and the trained NOx prediction model was evaluated using the evaluation data. As a result, it was possible to secure more than 93% prediction accuracy in the training data and evaluation data, and confirmed the validity of the proposed model. Based on the results of this study, it is judged that the research model can be applied not only to NOx but also to the artificial intelligence-based emission prediction model of other air pollutants with time-series characteristics.

REFERENCES

1. Lee, H., "Status and Improvement Measures for Total Air Pollution Load Management" National Assembly Research Service, 36(2019).
2. Karim, Z. A. A., Khan, M. Y., Rashid, A., Aziz, A. and Hagos, F. Y., "Attaining Simultaneous Reduction in NOx and SMOKE by Using Water-in-biodiesel Emulsion Fuels for Diesel Engine" A Journal of Engineering, 3, 1-21(2019).
3. Kohzadi, N., Boyd, M. S., Kermanshahi, B., and Kaastra, I. A comparison of artificial neural network and time series models for forecasting commodity prices, Neurocomputing, 10, 169–181(1996).
4. Ho, S. L., Xie, M., and Goh, T. N., A comparative study of neural network and Box-Jenkins ARIMA

- modeling in time series prediction, *Computers & Industrial Engineering*, 42, 371–375(2002).
5. Tabachnick, J., Fidell, B. G. and Ullman, L. S., *Using Multivariate Statistics*, Pearson Education, London (2007).
 6. Meyler, A., Kenny, G. and Quinn, T., “Forecasting Irish Inflation Using ARIMA Models” *Munich Personal RePEc Archive*, 11359(1998).
 7. Khashei, M., Bijari, M. and R aissi A rdali, G . A., “Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs),” *Neuro-computing*, 72(4-6), 956-967(2009).
 8. Zhang, G., Patuwo, B. E. and Hu, M. Y., “Forecasting with Artificial Neural Networks-the State of Art,” *Int. J. Forecast.*, 14(1), 35-62(1998)
 9. JCM Pires, FG Martins, SIV So usa, MCM Alvim-Ferraz and MC Pereira, “Prediction of the Daily Mean PM10 Concentrations Using Linear Models”, *American Journal of Environmental Sciences* 4 (5): 445-453(2008).
 10. [Yin Z hao, Y ahy a Abu Ha san, “ Machine learning algorithms for predicting roadside fine particulate matter concentration level in Hong Kong Central”, *Computational Ecology and Software*, 3(3): 61-73(2013).
 11. Hochreiter, S ., Schmidhuber, J ., Long s hort-term memory, *Neural Computation*, 9, 1735–1780(1997).
 12. Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., J iang, L . and C heng, Z., “ Time-series Well Performance Prediction Based on Long Short-Term Memory (LSTM) Neural Network Model” *J. Petrol. Sci. Eng.*, 186, 106682(2020).
 13. Li, Y. and C ao, H. , “ Prediction fo r T ourism F low based on LSTM Neural Network” *Procedia Comput. Sci.*, 129, 277-283(2018).
 14. Karakoyun, E. Ş. and Çıbıkdiken, A. O., “Comparison of ARIMA Time Series Model and LSTM Deep Learning Algorithm for Bitcoin Price Forecasting” *Proceedings of the Multidisciplinary Academic Conference*, 171-179(2018).
 15. K. G reff, R . K. Srivastava, J . K outnik, B. R. Steunebrink and J. Schmidhuber, “LSTM: a search space odyssey” *IEEE T rans. Neural N etworks and Learning S ystems*, vo l.28, no.10, p p.2222-2232(2017).
 16. H. Salehinejad, S. Sankar, J. Barfett, E.Colak and S. Valaee, “Recent advances in recurrent neural networks” online: arXiv preprint arXiv: 1801.01078(2018).
 17. A. Kri zhevsky, I. Sutskever, and G. E.Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proc. of the 25th International Conference on Neural Information Processing Systems*, Nevada, U.S.A., pp.1097-1105(2012).
 18. P. S ibi, S . A. J ones and P . S iddarth, “ Analysis of different activation function using back propagation neural networks” *Journal of Theoretical and Applied Information Technology (JATIT)*, vol.47, no .3, pp.1264-1268(2013).
 19. Z. Wang, “Mean squared error: love it or leave it? a new look at signal fidelity Measures,” *IEEE Signal Processing Magazine*, vol.26, no.1, pp.98-117(2009).