Ground Level Ozone Prediction for Delhi using LSTM-RNN

S. Geetha, L. Prasika

***Abstract:** Outdoor air pollutants are bringing adverse effects on the living being health. Air quality is deteriorating due to multi-pollutants such as sulfur dioxide (SO_2), Nitrogen dioxide (NO_2), Nitrogen oxide (NO_x), Ozone (O_3), Carbon Monoxide (CO), Particulate Matter 2.5 ($PM_{2.5}$), Particulate Matter 10 (PM_{10}), etc. Out of these multi-pollutants, ground level Ozone is creating major health issues in lungs, heart, etc. Ground level Ozone is formed due to reactions between Nitrogen, vehicle emissions, Industrial emissions, and gasoline with the presence of sunlight. Recently, Deep Learning Techniques are applied in all prediction problems. Here, we proposed the Recurrent Neural Network based LSTM prediction model to predict the ground level ozone. The model is created with the historical data collected from various stations in and around Delhi. The model is providing more accuracy to predict the ground level ozone than the stare-of-art techniques. The model is evaluated with normalized mean square error and mean absolute error.*

***Keywords:** Sulfur dioxide (SO_2), Nitrogen dioxide (NO_2), Nitrogen oxide (NO_x), Ozone (O_3), Carbon Monoxide (CO), Particulate Matter 2.5 ($PM_{2.5}$),*

I. INTRODUCTION

Air pollutants are classified into primary and secondary pollutants. Although the primary pollutants making portion of issues in the environment, the secondary pollutants are major influencer for health issues. Especially, the ground level Ozone called Tropospheric Ozone (O_3) involves in health and environmental issues. Mortality rate is swelling from 0.40 million to 1.23 million respiratory deaths with long term O_3 exposures [1]. World Health Organization (WHO) has set standards for concentrations of various air pollutants. For, ground level ozone (O3) is set as 100µg/$m^3$for 8 hrs. The ground-level ozone affects human health and vegetation [2]. Generally, ground level ozone is dominated by transportation [3]. So, there is a great need of developing prediction model to protect human, vegetation, materials, environment, etc. In this paper, we developed a prediction model using Recurrent Neural Network based Long-Short Term Memory (RNN-LSTM). RNN is behaving better in case of sequential dependencies are there in the dataset. Especially, the air quality time series data associated with sequential dependencies. However, many determined models are available to predict the tropospheric ozone, deep learning based models are ranging to offer enhanced predictions. The RNN-LSTM model is created to predict tropospheric ozone level with the air quality dataset collected from 4 stations in and around Delhi. Furthermore, the model shows better performance than the state-of-art techniques.

Related Work

Ground level O3 is predicted using a wide range of statistical techniques like Linear Regression [4], Principal Component Regression [5], Support Vector Regression [6], Fuzzy based Neural Network, etc.The prediction models are already existing with principal component regression for Delhi air quality data [7]. Forecasting ozone episodes using Neuro-Fuzzy approach for Delhi performs little better [8]. The accuracy obtained from these models are not satisfactory. Previously, the LSTM models are developed for traffic prediction, stock price prediction, even for few air pollutants. In this paper, the model is created using LSTM Network for predicting ground level Ozone(O_3).

II. METHODOLOGY

Initially, RNN is widely used in the air quality prediction. In RNN, there was a problem of storing previous data in memory cell. To overcome this problem, LSTM is emerged.

RNN-LSTM

RNN-LSTMNetworks contains 3 major gates. Input Gate, Output Gate, and Forget Gate [9]. LSTM framework is shown in the Fig.1. The maximum flow of values is controlled through input gate, how long the value exist in a memory cell is controlled by forget gate and output of the cell is controlled by output gate using activation function. In LSTM, Activation function is used to determine whether the value is in memory cell or forgotten [10] by using following formula.

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

Input gate determines which values used for updating and tanh is used to compute $\tilde{C}_t$. The formulas are

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C . [h_{t-1}, x_t] + b_C) \quad (3)$$

To update the old cell value, the following formula is used.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

Finally, to get the output of the model, the following formulas are used.

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

Revised Manuscript Received on December 28, 2018.

S. Geetha, Assistant Professor (Sr. Grade), Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India.

L. Prasika, Assistant Professor, Mepco Schlenk Engineering College, Sivakasi (Tamil Nadu), India.

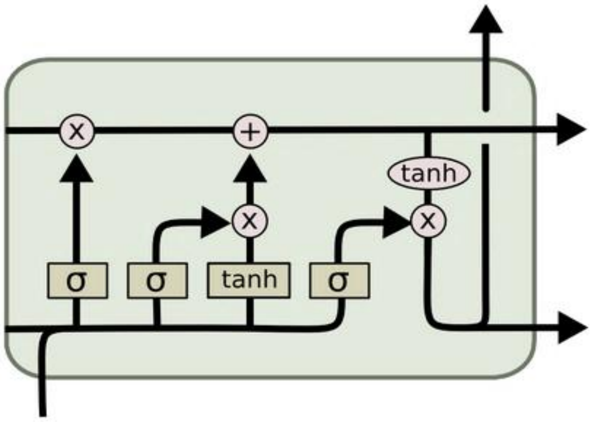

Fig.1: RNN-LSTM Framework

Here, x is the input, c is the cell output state, t is the time, h is the hidden layer output, i_t, f_t, and o_t are the outputs of the input gate,forget gate, and output gate. All three gates are using Sigmoid activation function.

Data Pre-processing

Central Pollution Control Board is monitoring air pollutants level through the various stations in India. In Delhi, the pollution level is moving beyond the tolerance. Especially, the ground level ozone became the primary reason for environment and health issues. So, the hourly based data is collected for Ozone (O_3) with other air pollutants from 4 stations (AnandVihar, Jagangirpuri, NehruNagar, Wazirpur) in and around Delhi for the period of Jan 2018 – Jun 2018. Dataset includes some missing values which are imputed by making mean of nearby values, when it is huge gap, those data are removed from the dataset.

Performance Measurement

Performance and accuracy of model is measured using Normalized Mean Square Error (NMSE), Mean Absolute Error (MAE). Equations for NMSE and MAE are as follows.

$$NMSE = \frac{\overline{(y_o - y_p)^2}}{\bar{y}_o \bar{y}_p} \qquad (7)$$

where y_o is the actual value, y_p is the predicted value and n is the total number of observations.

$$MAE = \frac{1}{n}\sum_{t=0}^{n}\left|y_o - y_p\right| \qquad (8)$$

where n is the total number of predictions, y is the predicted value.

The hourly averaged ground level ozone data of four stations is used for prediction through RNN-LSTM. And then, the performance is measured through Normalized Mean Square Error.

III. RESULTS AND DISCUSSION

Initially, the models are created with different approaches. The results from those models are compared to find the best model. The dataset is from 4 stations. Dataset is splitted into training set and testing set randomly. The model is trained with the training dataset and tested with the test dataset. The model is trained with various epochs and hidden layers finally the accuracy is arrived. The obtained training loss and testing loss of various station dataset is shown in Fig. 2, 3, 4, and 5.

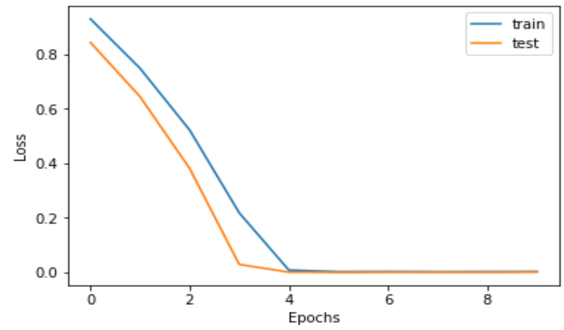

Fig. 2: Training and Testing Loss - Anand Vihar

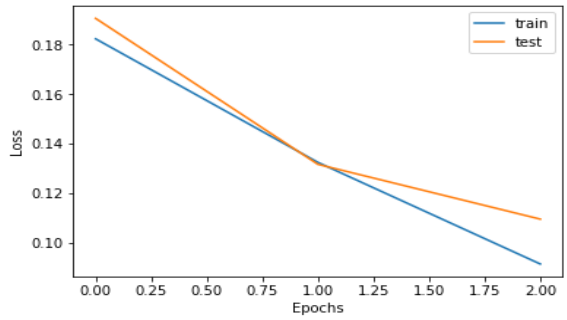

Fig.3: Training and Testing Loss - Nehru Nagar

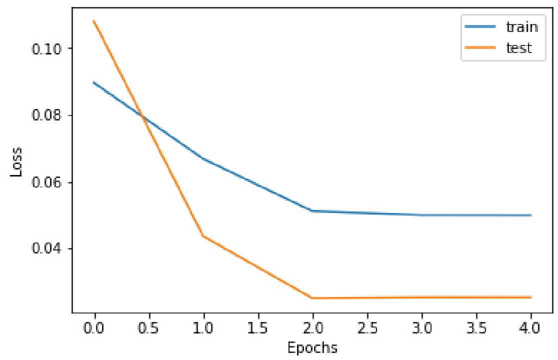

Fig.4: Training and Testing Loss - Jagangirpuri

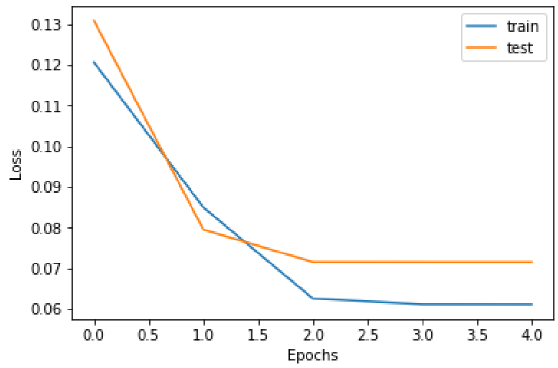

Fig.5: Training and Testing Loss - Wazirpur

O_3 dataset does not differ much in all four stations. So, the model's accuracy is achieved when the no. of epochs is 10, batch size is 10 and no. of neurons is 20 for all the stations.

The Table1 shows the performance evaluation for all four stations.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Station	Training		Testing	
	NMSE	MAE	NMSE	MAE
Anand Vihar	0.102	0.073	0.128	0.079
Nehru Nagar	0.128	0.082	0.179	0.110
Jagangirpuri	**0.089**	**0.050**	**0.065**	**0.025**
Wazirpur	0.096	0.061	0.107	0.071

Table 1: Performance Evaluation

The above table shows the normalized mean square error and mean absolute error. Out of 4 stations, the model shows better result for the dataset which is collected from Jagangirpuri with the normalize mean square error as 0.089 and mean absolute error as 0.050 for training set and 0.065 as normalised mean square error and 0.025 as mean absolute error for testing dataset. Next to that, the model performs well for Anand Vihar station dataset. The model shows better results compare to the state-of-art models.

IV. CONCLUSION

Air quality forecasting so important as it helps in protecting human health. The same way, prediction of ground level ozone is also very important. The developed model with LSTM is shows better prediction for the station Jagangirpuri than other stations. When the model is trained with the single station shows better performance than the training with all station's data. The missing values are imputed with simple mean which also affected the performance of the model. In future, the work will be extended to compare all the stations air quality data and also to include correlations of other air pollutants with the ground level ozone into the model.

REFERENCES

1. Malley, C. S., D. K. Henze, J. C. Kuylenstierna, H. W. Vallack, Y. Davila, S. C. Anenberg, M.C. Turner, and M. R. Ashmore, Updated global estimates of respiratory mortality in adults >30 years of age attributable to long-term ozone exposure. Environmental Health Perspectives 125 (8). doi: 10.1289/EHP1390, 2017.
2. Kim, S. E., & Kumar, A, Accounting seasonal nonstationarity in time series models for short-term ozone level forecast, Stochastic Environmental Research and Risk Assessment, 19(4), 241–248, 2005.
3. Andreja Stoji´c, DimitrijeMaleti´c, Svetlana Staniši´c Sto-ji´c, Zoran Miji´c, and Andrej Šoštari´c. Forecasting of vocemissions from traffic and industry using classificationand regression multivariate methods, Science of the TotalEnvironment, 521:19–26, 2015.
4. Sousa, S.I.V., Martins, F.G., AlvimTFerraz, M.C.M., Pereira, M.C., Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Environmental Modelling and Software 22, 97T103, 2007.
5. Al-Alawi, S.M., Abdul-Wahab, S.A., Bakheit,C.S., Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone, Environmental Modelling & Software 23, 396-403, 2008.
6. Petr Hájek and VladimírOlej. Ozone prediction on the basis of neural networks, support vector regression and methods with uncertainty, Ecological Informatics, 12:31–42, 2012.
7. Anikender Kumar, PramilaGoyal, Forecasting of air quality in Delhi using principal componentregression technique, Atmospheric Pollution Research, 436-444, 2011.
8. Dhirendra Mishra, P. Goyal, Neuro-Fuzzy approach to forecasting Ozone Episodes overthe urban area of Delhi, India, Environmental Technology & Innovation, 83–94, 2016.
9. Sepp Hochreiter and JurgenSchmidhuber, Long short-term memory. Neural computation, 9(8):1735–1780, 1997.
10. Xike Zhang, et al., A Novel Hybrid Data-Driven Model for Daily Land Surface Temperature Forecasting Using Long Short-Term Memory Neural Network Based on Ensemble Empirical Mode Decomposition, International Journal of Environmental Research and Public Health, 15, 1032, 2018.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*