VayuAnukulani: Adaptive Memory Networks for Air Pollution Forecasting

Divyam Madaan*
KAIST
dmadaan@kaist.ac.kr

Radhika Dua*
IIT Hyderabad
radhika@iith.ac.in

Prerana Mukherjee
IIIT Sricity
prerana.m@iiits.in

Brejesh Lall
IIT Delhi
brejesh@ee.iitd.ac.in

***Abstract*—Air pollution is the leading environmental health hazard globally due to various sources which include factory emissions, car exhaust, and cooking stoves. As a precautionary measure, air pollution forecast serves as the basis for taking effective pollution control measures, and accurate air pollution forecasting has become an important task. In this paper, we forecast fine-grained ambient air quality information for 5 prominent locations in Delhi based on the historical and real-time ambient air quality and meteorological data reported by Central Pollution Control board. We present VayuAnukulani system, a novel end-to-end solution to predict air quality for next 24 hours by estimating the concentration and level of different air pollutants including nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}) for Delhi. Extensive experiments on data sources obtained in Delhi demonstrate that the proposed adaptive attention based Bidirectional LSTM Network outperforms several baselines for classification and regression models. The accuracy of the proposed adaptive system is $\sim 15-20\%$ better than the same offline trained model. We compare the proposed methodology on several competing baselines, and show that the network outperforms conventional methods by $\sim 7-18\%$.** [1]

***Index Terms*—Air quality, Pollution forecasting, Real time air quality prediction, Deep Learning**

I. INTRODUCTION

Due to the rapid urbanization and industrialization, there is an increase in the number of vehicles, and the burning of fossil fuels, due to which the quality of air is degrading, which is a basic requirement for the survival of all lives on Earth. The rise in the pollution rate has affected people with serious health hazards such as heart disease, lung cancer and chronic respiratory infections like asthma, pneumonia, etc. Air pollution is one of the most important emerging environmental issues in the world. According to World Health Organization (WHO), in 2018, particularly India has 14 out of the 15 most polluted cities in the world. The increasing level of pollutants in ambient air in 2016-2018 has deteriorated the air quality of Delhi at an alarming rate. All these problems brought us to focus the current study on air quality in Delhi region. The establishment of a reasonable and accurate forecasting model is the basis for forecasting urban air pollution, which can inform government's policy-making bodies to perform traffic control when the air is polluted at critical levels and people's decision making like whether to exercise outdoors or which route to follow.

Prior work utilizes traditional handcrafted feature based approaches [1]–[4] and deep learning based methods [5]–[9] for forecasting expressway $PM_{2.5}$ concentration. However, these works do not take into consideration the crucial aspect of predicting the concentration levels of other pollutants which are equally important and have a serious impact on health conditions. Usually, the concentration of $PM_{2.5}$ is in the normal range but the concentration of other pollutants is very high. In [10], authors propose a framework for air quality estimation utilizing multi-source heterogeneous data collected from wireless sensor networks. However, it would still rely on the deployment of the sensor nodes. Therefore, in this paper, we aim towards addressing this gap and utilize a novel distributed architecture to simultaneously predict multiple pollutants present in the ambient air. In contrast to the aligning works [1]–[4], in this paper we capture the long-term temporal dependencies between the data sources collected using both direct as well as indirect sources and introduce the importance of attention based adaptive learning in the memory networks. Apart from that, we provide a standalone model to predict various pollutants ($PM_{2.5}, NO_2$ and PM_{10}) over a long period of time in future (next 24 hours).

The system we present, VayuAnukulani[2], predicts the air quality by estimating the concentration and level of different air pollutants. In this paper, we propose a novel approach to predict the concentration and pollution level i.e low, moderate and high pollution for different air pollutants including nitrogen dioxide (NO_2) and particulate matter ($PM_{2.5}$ and PM_{10}) considering air quality data, time and meteorological data of the past 24 hours. To the best of the author's knowledge, this work is one of the first holistic and empirical research where a single model predicts multiple pollutants and makes adaptive updates to improvise the future predictions.

The contribution of this work is threefold:

- The proposed approach uses heterogeneous urban data for capturing both direct (air pollutants) and indirect (meteorological data and time) factors for leveraging learning of the individual and holistic influences on air pollution using a novel distributed architecture.
- We build a novel end-to-end adaptive system that collects the pollution data over a period of 3 years for 5 locations

*Equal Contribution

[1] Code available at: github.com/divyam3897/VayuAnukulani

[2] The name is a direct Hindi translation for the two key words Air and Adaptatability - Air, translated as Vayu and Adaptatability, translated as Anukulani.

of Delhi, trains two models (forecast/classify) to predict the concentration and classify the pollution level for each pollutant ($PM_{2.5}$, PM_{10} and NO_2) for the next 24 hours, and then deploys a new model.
- We achieve a significant gain of $\sim 15-20\%$ accuracy with the proposed adaptive system as compared to the same offline trained model. The proposed system also outperforms several competing baselines and we achieve a performance boost of about $\sim 3-5\%$ in forecasting of pollution levels and concentration.

Some of the major problems that occur while implementing a prediction system to forecast the concentration and level of air pollutants are that air pollution varies with time and space, the data is often insufficient and inaccurate. Sometimes the concentration of pollutants is much beyond the permissible limits and contains several outliers. In order to forecast the concentration and level of air pollutants, it is necessary to collect data for every location independently as pollution varies drastically with geographical distance. In the proposed method, we address all the aforementioned challenges. We train separate models for each location as the pollution varies with varying geographical distance. Hourly air pollution and meteorological data are collected as pollution varies significantly with time. We use imputation techniques to deal with missing values and perform data preprocessing to select and use only the important features for air pollution prediction.

II. Proposed Methodology

Fig. 1 demonstrates the framework of the proposed solution. It consists of two parts: Offline Training and Online Inference.

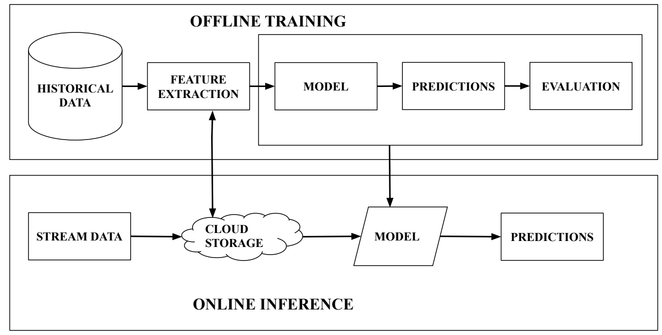

Fig. 1: Proposed framework of air pollution forecasting task

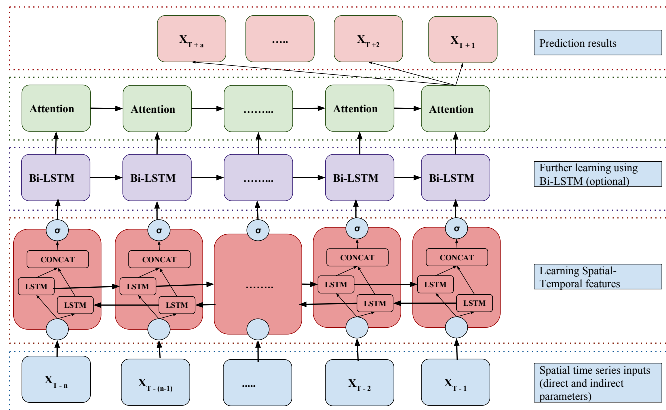

Fig. 2: BiLSTM-A consists of a BiLSTM layer and the attention mechanism layer. $\{x_{T-n}, x_{T-(n-1)}, ...x_{T-2}, x_{T-1}\}$ represent the historical input data sequence (direct parameters- air pollutants, indirect parameters- meteorological data and time) and $\{x_{T+a}, ...x_{T+2}, x_{T+1}\}$ represent the pollution forecasted values.

A. Offline Training

The offline training involves the collection of data followed by data preprocessing which involves the extraction of the important features for air pollution from heterogeneous data and handling the missing values in the collected data for various location. The model obtained is deployed and is used for predictions in the online inference.

1) Data Collection: The direct parameters consisting of all the air pollutants and indirect parameters which includes meteorological data and time from the **Central Pollution Control Board**[3] database which is publicly available and has been widely used for model evaluation. We utilize real-time air quality monitoring dataset collected for these locations to evaluate the performance of various models. The dataset contains the Air Quality Data, time and Meteorological data.

- **Air Quality Data:** The air pollutant variables in the air quality data are SO_2, NO_2, $PM_{2.5}$, PM_{10}, CO and O_3.
- **Meteorological Data:** The meteorological parameters include temperature, humidity, wind speed and barometric pressure.
- **Time:** Time includes hour of the day, seasons and month of the year.

All these variables of the past 24 hours are collected on hourly basis and extracted features from the collected dataset are used for evaluation of the models for predictions of the concentration of $PM_{2.5}$, PM_{10} and NO_2 after every 4 hours for the next 24 hours. Concentrations of all the pollutants are reported in $\mu g/m^3$.

2) Model and Evaluation: We propose an attention based BiLSTM [11] network called as BILSTM-A as shown in Figure 2. The input of the first BiLSTM layer is the historical data sequence followed by the attention module that assesses the importance of the representations of the encoded information and computes the weighted sum. Algorithm 1 outlines the proposed adaptive method. By running the algorithm every week on the hourly updated data for each location, the errors of the online adaptive algorithm are minimized.

B. Online Inference

The online inference in the framework is composed of the air quality data containing the air pollutant's concentration and meteorological data is collected real-time for every hour and is updated on the server. The cloud server plays the prominent role in updating the real-time data obtained every hour on the Cloud storage. After every week, the updated data along with the machine learning model is then passed by the server, where the training takes place. The trained model is then stored back

[3] http://cpcb.nic.in/

Algorithm 1 Algorithm for proposed adaptive method

Inputs: Data for each location $\{f_1, f_2, ..f_{n-1}, f_n\}$ and learning rate $\alpha = 10^{-3}$.
Initialize $F(x)$ = BiLSTM model with attention mechanism for N pollutants.
for $t \leftarrow 1$ ... T **do**
    Receive instance: x_t.
    Predict $\hat{y_t}$ for each pollutant for the next 24 hours.
    Receive the true pollutant value y_t.
    Suffer loss: $l_t(w_t)$ which is a convex loss function on both ${w_t}^T x$ and y_t.
    Update the prediction model w_t to w_{t+1}.
end for

on Cloud Storage. The stored model as part of the offline training process is then used to make the predictions every 4 hours for the next 24 hours and display it for the user to observe the forecasted values.

III. Evaluation and Analysis

In this section, we evaluate the efficacy and efficiency of the proposed architecture against several baseline models and analyze the experimental results. To our knowledge, there is no existing dataset for the current air prediction/forecasting techniques which includes different parameters, geographical locations and meteorological conditions which makes the comparison difficult. To have a fair comparison, we use the standard time-series prediction models as our baselines for all the tasks.

A. Baselines and our models

We evaluate our proposed method against many competing baselines. All methods and algorithms are implemented using Tensorflow [12] framework. We design the prediction model to forecast the continuous and categorical values for different pollutants. We split 80% of the dataset for training and 20% for test. All values are then normalized in the range [0, 1]. For training, batch size is set as 64 and we utilize Adam [13] optimizer with learning rate of 0.001 for gradient descent optimization. Dropout [14] is used as a regularization technique in all the models.

- **Random forest.** Baseline for prediction of pollutants concentration.
- **LSTM.** Base long short term memory model for pollution concentration and categories prediction.
- **LSTM-A.** Base long short term memory model with attention mechanism.
- **BILSTM.** Base bidirectional long short term memory model.
- **BILSTM-A.** Proposed bidirectional long short term memory model with attention mechanism for prediction of pollutants concentration.
- **CBILSTM-A.** Proposed bidirectional long short term memory model with attention mechanism for prediction of pollutants categories.
- **Initial BILSTM-A.** Base bidirectional long short term memory model with attention mechanism with initial dataset.
- **Adaptive BILSTM-A.** Proposed bidirectional long short term memory model with attention mechanism with updated dataset.

B. Pollution Concentration Prediction

Table I shows the $R^2, RMSE$ evaluation for different methods. It represents the evaluation for the next 4 hours. In general, we observe a performance boost with BiLSTM-A in the predictions as compared to the other baseline models by $\sim 7-18\%$. Random Forest performs well for the forecasting of $PM_{2.5}$ values. The proposed model of BiLSTM-A outperforms in the prediction of all the other pollutants. Fig. 3 shows the results of the predictions compared with the actual values for the BILSTM-A model for the next 4 hours. It shows that the proposed model achieves significant improvement in accuracy, especially in the scenarios of sudden change, which demonstrates that the proposed method indeed benefits the temporal predictions.

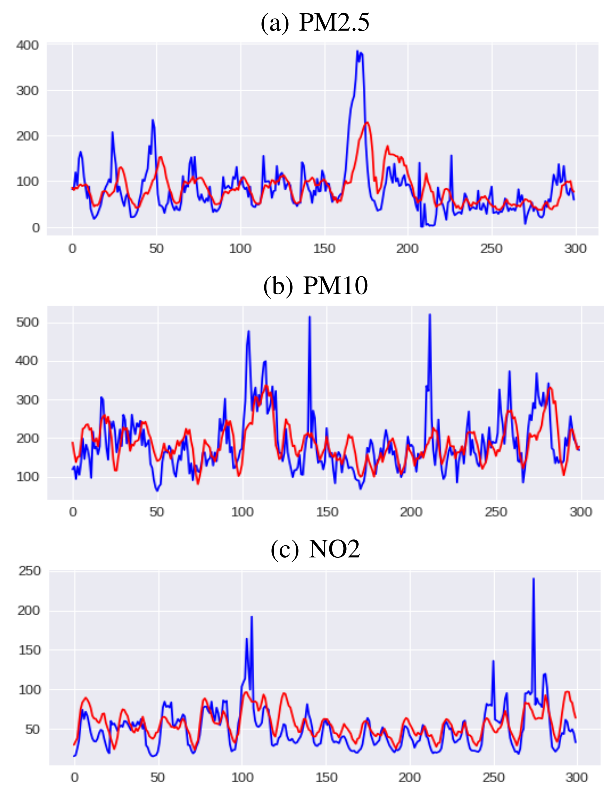

Fig. 3: Comparison between pollution estimates (red line) and actual measurements (blue line) for (a) $PM_{2.5}$ (b) PM_{10} (c) NO_2. The x-axis represents the number of samples and y-axis represents the concentrations of all the pollutants in $\mu g/m^3$.

C. Pollution Levels Prediction

For prediction for various pollutant levels, the concentration levels are divided into different standard threshold values by considering the critical values for each of them. $PM_{2.5}$ is divided into three categories: Low (0 - 60), Medium (60 -

TABLE I: Performance comparison of the proposed model with other baseline models for pollution values forecasting for future 4 hours on the basis of R-squared values and Root mean square error values. The highlighted values indicates the best performance.

Model	*Pollutants*	*R-square*	*RMSE*
Random Forest	$PM_{2.5}$	**0.35**	**40.69**
	NO_2	0.40	21.12
	PM_{10}	0.42	98.32
LSTM	$PM_{2.5}$	0.31	41.96
	NO_2	0.38	21.52
	PM_{10}	0.44	96.58
LSTM-A	$PM_{2.5}$	0.29	42.52
	NO_2	0.38	21.44
	PM_{10}	0.44	96.49
BILSTM	$PM_{2.5}$	0.30	42.07
	NO_2	0.38	21.47
	PM_{10}	0.44	96.77
BILSTM-A	$PM_{2.5}$	0.31	41.97
	NO_2	**0.41**	**21.08**
	PM_{10}	**0.45**	**96.22**

TABLE II: Performance comparison of the proposed model with other baseline models for pollution levels forecasting for future 4 hours on the basis of Accuracy, average precision and average recall. Higher values of accuracy, precision and recall indicates the better performance of the model. The highlighted values indicates the best performance.

Model	*Pollutants*	*Accuracy*	*Precision*	*Recall*
LSTM	$PM_{2.5}$	67.68	56.15	52.27
	NO_2	76.85	76.29	75.2
	PM_{10}	68.34	71.11	56.31
LSTM-A	$PM_{2.5}$	67.24	56.46	52.56
	NO_2	76.85	76.15	75.65
	PM_{10}	68.71	70.21	57.89
BILSTM	$PM_{2.5}$	67.96	58.35	53.12
	NO_2	77.32	76.75	75.86
	PM_{10}	**68.87**	**70.25**	58.36
BILSTM-A	$PM_{2.5}$	67.96	55.71	52.55
	NO_2	77.66	77.10	**76.26**
	PM_{10}	68.21	69.21	57.73
CBILSTM-A	$PM_{2.5}$	**70.68**	**61.06**	**55.8**
	NO_2	**77.88**	**77.56**	76.14
	PM_{10}	67.45	68.23	**58.52**

150) and High (150+), NO_2 into two categories: Low (0 - 50) and High (50+) and PM_{10} into three categories: Low (0 - 100), Medium (100 - 250) and High (250+) by considering the critical values for each of them.

Table II shows the evaluation of pollution level forecasting by comparing accuracy, precision, and recall of all the models. It represents the evaluation for the next 4 hours. For all the instances for the classification task we use the historical values of the last 24 hours. We categorize the output prediction by the various baseline regressor models and compare them with the output of the classifier model from our proposed method CBILSTM-A. It shows that BiLSTM-A shows a slight increase in accuracy compared to the other regressor models used in the air pollution levels prediction task and CBILSTM-A clearly outperforms all the baseline regressor models across all the evaluation metrics in the prediction of all the pollutant levels. This result confirms the effectiveness of a separate model for prediction of levels of various pollutants.

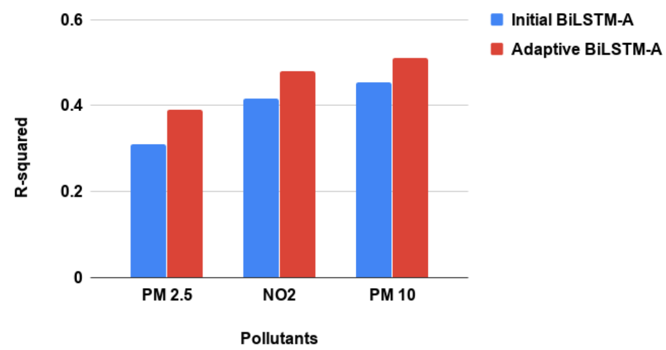

Fig. 4: Performance of Initial BiLSTM-A model with initial data compared to Adaptive BiLSTM-A model.

D. Adaptive Model Prediction

Due to the progressive temporal changes in the concentration of the pollutants, it is necessary to continuously update the collected data and the model, thus resulting in an adaptive model. We update the training data by collecting it from the Central Pollution Control Board every hour as explained in subsection II-B.

We compare the performance of the model on the initial collected data to the performance of the model which is updated every week on the real-time hourly updated data after a month of the collection of the initial data. We use our proposed BiLSTM-A model for evaluation of both the models. Figure 4 shows substantial improvement of the results after the real-time update which is indicative of the relevance of continuous updation of the model with the real time incoming data.

IV. Conclusion

Based on the historical and real-time ambient air quality and meteorological data of Delhi, we inferred the real-time and fine-grained ambient air quality information. We proposed a novel end-to-end system to predict the air quality of the next 24 hours by predicting the concentration of different air pollutants including nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}) for Delhi. The results showed the performance boost with the proposed method over other well-known methods for regression models. In future work, we intend to explore more powerful modeling techniques along with the traffic density data, as a way to model the traffic density of the monitored location to get better results.

V. Acknowledgments

We would like to thank Dr. Aakanksha Chowdhery, Google Brain for the inspiring guidance and constant encouragement during the course of this work. We acknowledge the Marconi Society, IIT Delhi, and Celestini Program India to support the project. We owe our heartfelt gratitude to them for providing us an opportunity to pursue work with excellent laboratory facilities at IIT Delhi. We would also like to thank Central pollution control board for the dataset which has played a very important role in the completion of this work.

References

[1] Y. K. D. H. S. E. D. K. Ming Dong, Dong Yang, "Pm2.5 concentration prediction using hidden semi-markov model-based times series data mining," vol. 36, no. 5, pp. 9046–9055, 2009. [Online]. Available: https://doi.org/10.1016/j.eswa.2008.12.017

[2] J. Sun, Wei; Sun, "Daily pm2.5 concentration prediction based on principal component analysis and lssvm optimized by cuckoo search algorithm," 03 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301479716309835

[3] S. Zhang, X. Li, Y. Li, and J. Mei, "Prediction of urban pm 2.5 concentration based on wavelet neural network," in *2018 Chinese Control And Decision Conference (CCDC)*. IEEE, 2018, pp. 5514–5519.

[4] Z. Shang, T. Deng, J. He, and X. Duan, "A novel model for hourly pm2. 5 concentration prediction based on cart and eelm," *Science of The Total Environment*, vol. 651, pp. 3043–3052, 2019.

[5] V. O. Li, J. C. Lam, Y. Chen, and J. Gu, "Deep learning model to estimate air pollution using m-bp to fill in missing proxy urban data," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[6] Y. Pengfei, H. Juanjuan, L. Xiaoming, and Z. Kai, "Industrial air pollution prediction using deep neural network," in *International Conference on Bio-Inspired Computing: Theories and Applications*. Springer, 2018, pp. 173–185.

[7] İ. Kök, M. U. Şimşek, and S. Özdemir, "A deep learning model for air quality prediction in smart cities," in *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1983–1990.

[8] T. Li, H. Shen, Q. Yuan, X. Zhang, and L. Zhang, "Estimating ground-level pm2. 5 by fusing satellite and station observations: A geo-intelligent deep learning approach," *Geophysical Research Letters*, vol. 44, no. 23, 2017.

[9] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. M. Zhang, "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[10] C. Feng, W. Wang, Y. Tian, X. Que, and X. Gong, "Air quality estimation based on multi-source heterogeneous data from wireless sensor networks," in *Wireless Communications and Networking Conference (WCNC), 2018 IEEE*. IEEE, 2018, pp. 1–6.

[11] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6645–6649.

[12] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *CoRR*, vol. abs/1603.04467, 2016. [Online]. Available: http://arxiv.org/abs/1603.04467

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html