Exploring Transformers for Urban Air Quality Inference

Hitarth Gandhi 19110087 **Ishan Prayagi** 19110194 **Somesh Pratap Singh** 19110206 **Vishal Soni** 19110207 **Xhitij Choudhary** 19110044

Mentors: Harsh Patel, Palak Purohit

Abstract

Growing air pollution and related environmental and health hazards make the problem of Air Quality Inference of utmost importance. Facilities are set up in several regions to infer the said air quality data. However, they are limited in number over the vast expanses of countries, given the cost limitations in setting them up. Hence, we need to develop a method to predict air quality data of regions where AQI inference stations are not available. To solve this problem, we propose a complex neural approach by combining Transformers, Fully Connected Neural Networks (FNNs), and Graph Neural Networks (GNNs). The Transformer layer will capture long-range dependencies among the temporal features. The GNN layer will extract information and dependencies from the spatial features. A FNN will combine the outputs from Transformers and GNN to find dependencies between spatial and temporal features.

1 Introduction

Air pollution is a serious environmental health hazard. Almost 9 out of 10 people living in urban areas are affected by air pollution. It can be detrimental to lung development and lead to other respiratory diseases like asthma, emphysema, and chronic obstructive pulmonary disease (COPD).[1]

The air quality index (AQI) is widely used to measure air quality. For a specific air pollutant, its individual air quality index (IAQI) in an area is measured by a monitoring station, reflecting the real-time concentration of the pollutant. AQI is the highest IAQI values among all kinds of air pollutants.

The Air Quality Index (AQI) of several places in India has crossed 400, whereas safe levels are in the range 0-100.[2,3]

The cities, which are heavily polluted, have lesser number of air quality inference stations.[4] These areas have high populations in the range of 0.5 to 2 million which are affected by this perilous pollution.[5]

The deployment of more stations is one visible solution to the problem. However, due to several limitations, there is a plateau in the number of stations that can be set up. Thus, it becomes essential to develop solutions to predict AQI data of regions with sparse AQI inference stations deployment. This will help residents take necessary steps in a timely manner to stay safe if the levels were to exceed safe levels.

Personal Motivation

Air pollution is a global problem that is increasing by the day. We find it in our keen interest to deploy our skills for contributing to solving this global problem. The novelty of using Transformers to solve this problem instead of LSTMs is what inspires us. We are also planning on exploring potential research publications and conference presentations as well as making a submission at the NASA Airathon[6] to make an impact with our model.

2 Related Work

Past approaches include using classical emission models to simulate the flow of air particles based on numerous empirical assumptions and parameters. These works include Gaussian Plume Modeling(Wahab et al., 2014) and Computational Fluid

[1] https://www.niehs.nih.gov/health/topics/agents/air-pollution/index.cfm

[2] https://www.airnow.gov/aqi/aqi-basics/

[3] https://theprint.in/india/delhi-records-worst-air-quality-in-five-years-a-day-after-diwali/761984/

[4] http://cpcbenvis.nic.in/airpollution/monetoring.htm

[5] https://www.census2011.co.in/census/city/41-hisar.html

[6] https://www.drivendata.org/competitions/88/competition-air-quality-pm/page/424/

Dynamics(Scaar et al., 2013). These models utilize many ideal world approximations and assumptions which are not true in the real world. Thus, they give very unsatisfactory results.

Other approaches include Statistical modeling such as mathematical formulations, regression models, and Neural networks. These methods include

- K-nearest neighbors, where we select k nearest stations and then average the AQI values from those stations.
- Linear Interpolation, where we take a weighted average of IAQI values from all the stations. The weights assigned to each station are inversely proportional to the distance of the station.
- Gaussian Interpolation. Interpolation based on Gaussian distribution.
- Gaussian Process Regression, This is a non-parametric Bayesian regression model.
- Feedforward Neural Networks, a simple Neural Network with dropout and L2 normalization, is used to predict AQI values. For sequential features, only the latest values are used.
- Support vector regression and Long short-term memory based on Deep Learning are used to classify the AQI values.(Janarthanan et al., 2021)
- U-Air(Zheng et al., 2013), a classification model which contains two classifiers each for spatial features and temporal features. Then the model combines output from both classifiers to infer AQI. But this separation of features may cause the model to overlook dependencies between spatial and temporal features.
- ADAIN(Cheng et al., 2018), an RNN-FNN hybrid model which contains two groups of input layers each for spatial and temporal features. It also contains an attention layer to learn the importance of different stations automatically over using the k-nearest neighbors method for the same like in some previous papers. This is currently the state-of-the-art model in predicting the AQI.

There are several other works that solve similar problems like predicting the flow of crowd in a city(Zhang et al., 2016) and real-time public transportation-based crowd prediction systems(Liang et al., 2016). But these works can not be extended to predicting AQI due to the fine-grained nature of air particles.

The results of the state of the art model ADAIN are described in the below figures:

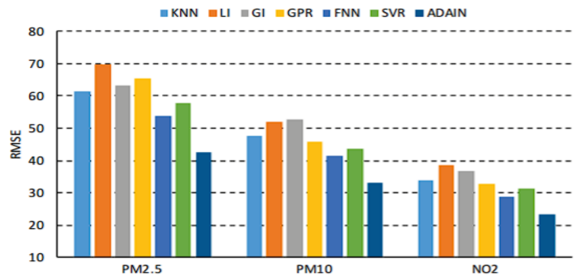

Figure 1: ADAIN v.s. competing regression methods

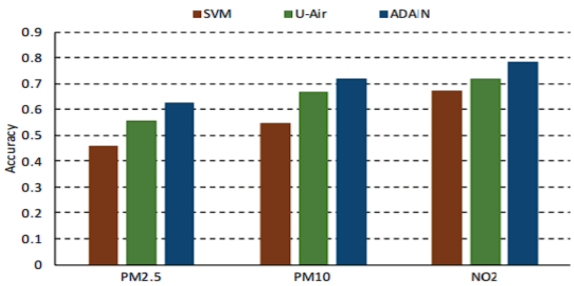

Figure 2: ADAIN v.s. competing classification methods

3 Dataset

Our model will be trained on the Beijing Multi-Site Air-Quality Data Data Set[7]. We will compare our model on two well-known and verified datasets: Beijing Multi-Site Air-Quality Data Set and NASA Meteorological Data[8] (available through a public sp3 bucket). Both of these datasets are official and authenticated.

4 Methodology

Since the state-of-the-art model, ADAIN, is not open-source, we will first implement the ADAIN model itself. We will then implement the other baseline models mentioned above, such as Linear

[7]https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data

[8]https://www.drivendata.org/competitions/88/competition-air-quality-pm/data/

Interpolation, KNN model, and Support Vector Regression.

Then we will experiment on the ADAIN model by making changes in different layers of the model.

4.1 Our Approach

We plan to use Transformers instead of LSTMs in the ADAIN model proposed by Cheng. The existing methods have a weak ability to capture long-term dependencies and complex relationships from time-series data. Recent studies have proven that Transformers with attention mechanisms are far better learners of long-range dependencies than LSTMs in machine translation tasks in NLP, object detections, and classification(Zhang et al., 2021). Thus we propose to further improve the ADAIN model by using Transformer layers instead of LSTM layers for temporal data.

Further, we can extend it by implementing a Graph Neural Network (GNN) for spatial features instead of a superficial Fully Connected Layer in the ADAIN model. GNNs are a generalization of conventional Convolutional Neural Networks. GNNs can be used to deal with homogeneous graphs. Thus we can use them to extract information from spatial features and find dependencies among them.

4.2 Model Architecture

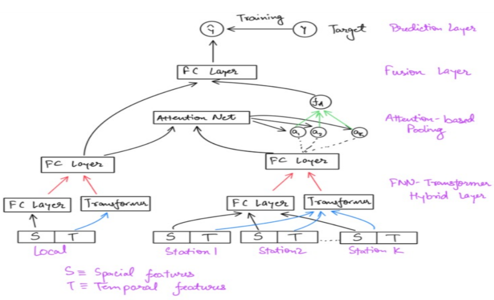

Figure 3: Model Architecture

The model input layer consists of two groups of features - local features for the required location and station-oriented features for each station.

The next layer is an FNN-Transformer hybrid layer. The FNN layer is for spatially related features. It learns spatial dependency among those features. The Transformer layer is for temporally related features that aim to learn temporal dependency among them.

Further, another fully connected layer is added to capture dependencies among the spatial and temporal features.

The next layer is the Attention layer. This automatically learns weights for each station. A multi-layer perceptron-based parametrized attention score is used to assign weights. The attention scores are normalized using a softmax function.

Then a Fusion Layer is added to learn the dependency between the local features and stations.

The last layer is the prediction layer, where the output of the fusion layer is transformed to get the IAQI value for the target location.

4.3 Dataset Preprocessing

- Handle null values in the dataset by experimenting with various interpolation techniques.
- Pre Processing the dataset to match it to the specifications of the model.

5 Evaluation

The obtained IAQI values will be examined, and the performance of various regression approaches will be evaluated by the commonly used metric Root Mean Squared Error (RMSE). For the classification models with discrete outputs, the output would be converted into the corresponding IAQI values, and then they would be evaluated by the accuracy metric. Here accuracy would mean the correct estimation of the test cases to the total number of cases. At the end of the project, a lower RMSE or a higher accuracy of outputs than the state-of-the-art approaches would be desired.

6 Project Management

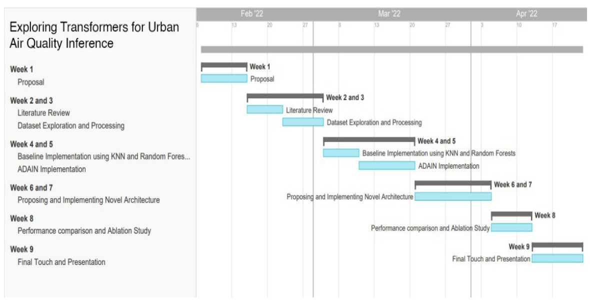

Figure 4: Gantt Chart for the Project

References

Weiyu Cheng, Yanyan Shen, Yanmin Zhu, and Linpeng Huang. 2018. A neural attention model for urban air quality inference: Learning the weights of monitoring stations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi. 2021. A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities and Society*, 67:102720.

Victor Liang, Richard Ma, Wee Siong Ng, Li Wang, Marianne Winslett, Huayu Wu, Shanshan Ying, and Zhenjie Zhang. 2016. Mercury: Metro density prediction with recurrent neural network on streaming cdr data. pages 1374–1377.

H. Scaar, T. Teodorov, T. Ziegler, and J. Mellmann. 2013. Computational fluid dynamics (cfd) analysis of air flow uniformity in a fixed-bed dryer for medicinal plants. *Acta Horticulturae*, (1008):119–126.

Magdy Wahab, Khaled Essa, H.M. Elsman, A. Soliman, S.M. Elgmmal, and Ali Wheida. 2014. Derivation of the gaussian plume model in three dimensions. *Mausam*, 65:83–92.

Junbo Zhang, Yu Zheng, and Dekang Qi. 2016. Deep spatio-temporal residual networks for citywide crowd flows prediction.

Zhen Zhang, Shiqing Zhang, Xiaoming Zhao, Linjian Chen, and Jun Yao. 2021. Temporal difference-based graph transformer networks for air quality pm2.5 prediction: A case study in china.

Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, page 1436–1444, New York, NY, USA. Association for Computing Machinery.