PREDICTING AND ANALYZING AIR QUALITY FEATURES EFFECTIVELY USING A HYBRID MACHINE LEARNING MODEL

[1*]Ms PAGADALA CHAITANYA, Mr. KASHIMALLA NARESH[2],
Mr. M. VISHNU VARDHAN REDDY[3]
[1,2,3]Assistant Professor, Dept Of CSE ,
St.Martin's Engineering College, Dulapally, Kompally, Secunderabad, Telangana.

Abstract:

The interpolation, prediction, and feature analysis of fine-gained air quality are three essential subjects in the domain of urban air computing. As a result, the answers to these questions may have significant social and technological ramifications in the fight against air pollution. Most of the previous work handles the three challenges individually via distinct models. In this research, we offer a broad and effective strategy to handle the three issues in one model dubbed the Hybrid Machine Learning Model.. The core principle of HMLM resides in integrating feature selection and learning in multiple levels of the deep learning network. Using unlabeled spatio-temporal data, the proposed technique uses feature selection and association analysis to uncover the most significant characteristics for air quality change and improves interpolation and prediction performance. We test our technique with numerous experiments based on genuine data sources gathered in Delhi, India. Experiments reveal that HMLM is superior to the peer models from the current literature when addressing the themes of interpolation, prediction, and feature analysis of fine-gained air quality.

Key words : Keywords: air pollution prediction; deep learning; Hybrid Machine Learning Model;

I. INTRODUCTION:

In 2019, as part of a global assessment, it was determined that 21 out of the 30 most polluted cities were in India. As a result, according to data from iqair.com, India rose to the 5th position in the world. The US AQI rating averaged out at 152 and the PM2.5 level measured was 58.08µg/m³. This concentration was 5 times greater than that advised by the World Health Organisation (WHO) (WHO). This is an overall improvement over the 2018 value of 72.54µg/m³. This means serious health problems for most of the country.

Per 50 over cent of this pollution originates from industry, followed by 27 per cent from automobiles, 17 per cent from agricultural burning and 7 per cent from residential cooking. Over 2 million Indians lose their lives to reasons attributable to air pollution.

"In the metropolitan regions, most of the pollution comes from industry and cars, but in the rural areas, most comes from the burning of organic debris. This substance is utilised as a fuel for the residential stoves, and also in the heaters required to keep the dwellings warm in the winter months. During fall and winter, vast quantities of stubble are burned in the fields as a manner of preparing the land for the following crop. The alternative of re-plowing the leftovers into the soil is substantially more expensive than using this technique. This is especially problematic since many people dispose of their rubbish in open flames"[1].

This combines with other pollutants to rank India as the world's third highest emitter of greenhouse gases, after China and the USA. Because of a lack of strict enforcement, the Air (Prevention and Control of Pollution) Act, which was passed in the early 1980s, has had little effect.

Official attitudes to the bad air quality is shifting, particularly after schools in Delhi had to shut for many days in December 2017 owing to the unsafe amounts of pollutants present in the air. As more people are aware of the implications of inhaling poor quality air, the pressure is rising on the government to do something about it.

SOURCES OF INDIA'S POLLUTED AIR

Fuel used for residential stoves is commonly prepared from a moist combination of bits of wood, dried leaves, hay and dried animal manure. To dry, they are shaped into discs and placed in the sun. When burned in stoves or chullas, it creates five times as much smoke and other pollutants than coal does. It is anticipated that in excess of 100 million homes use these stoves up to 3 times a day, 7 days a week. For the most part, isolated places lack access to reliable sources of electricity or other environmentally friendly energy sources. The usage of these sorts of stoves is widespread, even in places where power is readily accessible, and it is responsible for 24% of the city's pollution.

"Some Indian auto-rickshaws and taxis operate on petrol that has been contaminated with other, cheaper substances. This is a regular occurrence across all of South Asia. Fuel taxes in India are substantially higher on gasoline than on diesel, thus aggravating an already precarious economic condition in the system. This, in turn, has a greater level than kerosene since kerosene is designed to be used as a cooking fuel. Other volatile liquids such as lubricants and solvents carry little or no tax and hence make suitable additives to blend with the higher-priced fuels. Over the course of a month, this falsification may save as much as 30% for someone on a modest income.

The sheer number of vehicles attempting to squeeze onto the few accessible roadways in India's major cities and towns is a major source of traffic congestion. A lack of split highways inside cities and traffic accidents caused by India's haphazard road conditions and lax enforcement of the country's regulations are further factors in the situation. Traffic is slowed to a crawl as a result of the snarls caused by intersections. Observation stations at major junctions report much larger numbers than those found elsewhere"[2].

Dust created by the demolition and subsequent development of new houses adds to the bad quality of air in the city. Desert dust is carried into cities by winds that reduce in pressure when they pass over high-rises and other structures.

The air quality in the capital of Delhi always slips to the “severe” level throughout the winter months. An important contributing factor is that harvest stubble is burned to make way for next season's crops, which is a common practise. An estimated 32% of Delhi's PM2.5 pollution is attributed to this one source. At 292micrograms per cubic metre, the amount is 5 times higher than the World Health Organisation’s recommended safe level. Wind and rain disperse airborne particles, and the weather conditions influence these processes significantly.

Another important contributor to the air pollution was the Badarpur Thermal Power Station. This was erected in 1973 and provided a modest 8 per cent of Delhi’s power but was answerable for 80-90 per cent of the particulate matter. As a result of "The Great Smog of Delhi," which occurred in November 2017, the plant was forced to temporarily shut down, however it was reopened in February 2018. Due to the quantity of pollutants it produced, however, it was permanently shut down by the end of 2018.

EFFECTS DOES THIS POOR AIR QUALITY HAVE ON HEALTH

“The levels of the pollutant PM 2.5 are typically much over the World Health Organisation’s recommended threshold of exposure (sometimes over 5 times higher) and this leads to major respiratory difficulties for individuals exposed to it. This comes from both outdoor and domestic air pollution. Poor air quality was blamed for the deaths of almost 1.6 million people in 2019, according to official statistics. The reason of death varied from strokes, diabetes, lung cancer and myocardial infarctions. According to the State of Global Air 2020 report, which was released earlier this year, air pollution is now the leading cause of mortality worldwide”[3].

More than 100,000 newborns died in the first month of their lives as a result of exposure to household and outdoor particulate matter pollution. A substantial number of these fatalities were connected to the usage of solid fuel biomass (charcoal, wood and dried dung cakes) used for cooking and heating dwellings.

Comparatively speaking, Indians are exposed to 83.2 g/cubic metre on average of PM2.5 pollutants, as opposed to the 8 g/cubic metre average in cleaner nations.

Because PM 2.5 particles penetrate the alveoli of the lungs, poor air quality has a significant impact on the human respiratory system. They may then go through the body's tissues and potentially reach the heart from this point. Reduced lung capacity, sore throats, coughing, weariness, lung cancer and headaches are all frequent signs of exposure to polluted air.

According to a doctor at Delhi's Sir Ganga Ram Hospital, most of his lung cancer patients in their 60s were male smokers when he first began practising medicine 30 years ago. But lately the doctor has observed that his patients now are mainly non-smokers, and around 40 per cent are female. Patients are also much younger today, with 10% or so in their 30s and 40s, he noted. Even in the lungs of young patients, black deposits may be seen which would have been inconceivable 30 years ago. COPD or chronic obstructive pulmonary disease is currently the second-largest cause of mortality after heart disease.

It has been reported that the levels of PM2.5 particles are impacted by the population density

India’s air quality be improved

On January 1st, 2019, India's government officially launched the National Clean Air Program (NCAP) to combat the issue. More than 122 of the world's most polluted cities will see a 20-30 percent reduction in pollution levels by 2024. Actions being done in New Delhi, Ahmedabad and Pune include the development of health risk messaging programmes, the increase in the number of monitoring stations and improved regulation of industrial pollutants.

Sulfur dioxide (SO2), nitrogen oxides (NO2 and NO), PM10 particles, and suspended particulate matter are also commonly monitored by NCAP (SMP). These will be monitored by 308 stations spread throughout 115 towns and cities in 25 different states and four other territories across the world.. Meteorological measurements are also recorded such as wind speed and direction, relative humidity and temperature. Particulate matter and gaseous contaminants are measured on a regular basis. These measurements are collected twice a week and will provide 104 observations over a period of 1 year.

PM10 particles have declined in some places, notably Solapur and Ahmedabad, in recent years. Local authorities and the reduction of sulphur in diesel fuel are likely to be responsible for this.

Sulphur Dioxide levels are dropping in residential areas of Delhi. These last few years have seen a resurgence in Mumbai, Lucknow, and Bophal. Liquid Natural Gas's growing availability and popularity have been credited with this decrease in emissions (LNG). To replace biomass in home cooking stoves and auto-rickshaws, this fuel is being promoted.

The administration in New Delhi established the “odd/even” regulation in late 2017. This simply implies that automobiles with a registration plate ending in an even number are barred from the city core on certain days of the week. For various days, the odd numbers follow the same pattern.

India’s aims over the next several years to minimise air pollution include the adoption of over 1,000 electrically powered buses and the upgrading of engines utilising fossil fuels to meet the demanding BS6 requirements. All power plants in India are expected to use renewable energy by 2023, when it is predicted that 25% of all privately owned cars would be electric vehicles (EVs).

Vehicles that are 15 years old or older, or that do not meet the BS6 emission regulations, are prohibited from operating on municipal streets.

The burning of straw at the conclusion of each harvest is being discouraged in rural areas by promoting the use of machines that turn organic waste into fertiliser.

The public will be better informed about potential changes in air quality thanks to the emergence of new technologies.

It is anticipated that the decrease in carbon emissions would lower CO2 emissions by 20% by 2030 and eliminate them entirely by 2075.

“One of the ideas being investigated is the planting of 1.35 billion native trees over the next 10 years along a 1,600-kilometer "green" corridor from Gujarat to Delhi in order to organically purify the air”[10].

“Delhi is nearly free from the usage of kerosene as a fuel and approximately 90 per cent of the people now use LPG (NPG) for cooking. Only 10% of the world's population still uses conventional fuels including wood, coal, and animal dung”[9].

It is anticipated that all of Delhi's energy needs would be met by clean, renewable sources of energy by the year 2021.

CLEANEST AIR IN INDIA

“The cleanest city in India is Satna in the state of Madhya Pradash with a 2019 PM2.5 measurement of 15.5µg/m³ and a US AQI value of 58. The second cleanest city was Kumbhori in the state of Maharashtra with a result of 20.3µg/m³. Compare these numbers to the worst air in Uttar Pradesh, where Ghaziabad has an AQI of 179 and a reading of 115 ng/m3, and to the second dirtiest city in the country, Delhi, which has a level of 100 ng/m3.In some places of India, air quality has been getting better since 2018, although very modestly. The number of days when the pollutants were above acceptable values for PM2.5 reduced in 2019 by comparison”[8].

“The geology of Northern India and its closeness to the Himalayas means that it is exceedingly difficult for filthy air to escape. During the winter, when the wind is less, the region resembles a bowl into which pollutants cannot be expelled.The interpolation, prediction, and feature analysis of fine-gained air quality are three essential subjects in the domain of urban air computing. As a result, the answers to these questions may have significant social and technological ramifications in the fight against air pollution. Most of the previous work handles the three challenges individually via distinct models“[4].

II. RELATED WORK

Air Quality Prediction

“The machine learning models for air quality prediction may be classified into two categories: basic learning models and deep learning models. Basic learning models include linear regression, supporting vector regression, random forest, and LightGBM. Land Use Regression (LUR) generates air quality forecasts by a linear regression model that takes into consideration several aspects such regional population level, transportation condition, and land use condition. LUR does not incorporate the intricate spatiotemporal connection of air pollution data, hence the accuracy of prediction is weak. Later, the fundamental time series model autoregressive integrated moving average model (ARIMA) arose, which was utilised for time series forecasting with high periodicity. However, it does not function well for difficult weather situations. Random forest , LightGBM deep learning algorithms have been commonly used approaches in air pollution prediction”[5]. Later, in order to further improve the accuracy of prediction, Zheng et al. proposed U-Air, which uses a spatial classifier based on an artificial neural network (ANN) and a temporal classifier based on the linear-chain conditional random field (CRF) to capture temporal and spatial characteristics.” Convolutional neural networks (CNN) are used to handle data from Euclidean structures. For example, they are particularly successful in the area of image identification, and it is difficult to use CNN directly to capture the spatial connections between monitoring stations for sparse graph topologies consisting of monitoring stations. It is possible to use the ConvLSTM model introduced in to define the spatiotemporal interaction between monitoring stations in Euclidean space by combining CNN and LSTM. The advent of Graph Convolutional Networks (GCN) has made up for the inadequacies of CNN and is commonly employed in traffic data. GCN has accomplished the full usage of the traffic network. GAT [9] are suggested on the basis of GCN, utilising an attention mechanism, and are excellent at capturing dynamic interactions between nodes. The ST-GAT model introduced by Zhang et al. can dynamically capture the dynamic dependencies in the traffic network, making the traffic speed prediction results more advanced than current models"[6].

"Through the introduction of the above connected studies, it can be shown that the present air quality prediction algorithms seldom explore the spatial interaction between many monitoring stations. A few spatiotemporal prediction models lack the capacity to dynamically represent spatial correlation depending on weather and other associated elements. The only approaches that can dynamically simulate spatial correlation do not include how to cope with limited training data. Some existing methods in the area of transfer learning and metalearning can solve the insufficient-training-data situation to a certain extent by transferring the knowledge from other source domains, but these methods lack the ability to adapt to the air quality spatiotemporal prediction models and cannot be directly applied to the scenarios targeted in this article. For this reason, this work provides a spatiotemporal model for air quality prediction and a metalearning technique for this model. The prediction model can dynamically and precisely simulate the temporal and geographical connection in air quality prediction. The metalearning technique is used to construct a more accurate prediction model in the situation of limited training data. As far as we know, it is the first time that metalearning has been utilised for air quality prediction" [7].

Urban air computing has various difficulties, however, since the underlying data is unique. First, as there are inadequate air-quality-monitor stations in a city owing to the high cost of installing and maintaining such a station, it is costly to get labelled Training samples when dealing with fine-gained air quality.

There is a technical issue with the air quality monitors that is causing the incorrect labelling of the devices. In usually, each station only has one monitor device which has to be maintained at intervals, so there will be no outputs for the station while the device is being serviced, recalibrated, or has other difficulties.

III. PROPOSED SYSTEM:

In this we suggested Hybrid Machine Learning Model (ISOA AND SVR) (ISOA AND SVR) This study is driven to handle all these issues by exploiting the information included in the unlabeled data and the spatio-temporal data, and conducting feature selection and association analysis for the urban air related data. Obtaining huge numbers of unlabeled instances is far more cost-effective than doing so for a small number of annotated ones.

In this work, the ISOA method was utilised to optimise the penalty parameter c, width parameter g and loss parameter p in SVR. The mean square error (MSE) is the fitness function of ISOA algorithm. The fitness function of the kth training sample is determined by the following Equation:

$$\boldsymbol{MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \dot{y}_i)}\text{------------}\mathbf{(1)}$$

where n is the number of samples and yˆi and yi are the actual and the forecasted return AQI value, respectively.

The algorithm flow of the proposed Hybrid Machine Learning Model (ISOA-SVR prediction model) is as follows:

Step 1: Set the control parameters of the proposed ISOA algorithm.

Step 2: Initial population P.

Step 3: Map P into the c, g and p of the SVR and calculate its fitness evaluation f by use of Equation (1), then f itpbest, f itgbest and avg f it are calculated.

Step 4: Update weight parameters ω by Equation (2).

$$\omega = \int_{\omega max}^{\omega max-(\omega max-\omega min)\left(\frac{fitpbest-fitgbest}{avgfir-fitgbest}\right)T},\text{--------(2)}$$

Step 5: Update C by Equations (3) and (4).

A = a − a(t/T) (3)

C = AP (4)

Step 6: Update M by Equations (5).

B = e A brand ---------------------------------(5)

Step 7: Update D by Equation (6).

D = |C + M| ---------------------------------(6)

Step 8: Update attack positions P .

Step 9: The fitness values f are sorted from small to large. Update the attack position P .

Step 10: If the stopping criterion is met, then go to Step 11. Otherwise, go to Step 3.

Step 11: Output the best positions that is mapped into the c, g and p of the SVR. Then train and test the SVR.

However, their AQI differencenis rather large, where the first picture belongs to the class AQI-1, and the second—to the class AQI. The meteorological circumstances make a big impact in classifying, therefore the requirement for incorporating them into the classifier, as we have done in our own pretrained inception network.

Considering the difficulty provided by visually identical photos with drastically varied AQIs, the proposed model's worth is underlined considering that it classifies images with 0.896 accuracy on the training set and 0.763 accuracy on the testing set. Using just photos and meteorological information, our approach enables us virtually ideally to determine if pollution is present.

Conclusion

Air pollution prediction is an important analytical challenge with the ability to give decision support skills to solve air quality concerns. In order to do this, we looked at the use of deep learning architectures in this article. HYBRID MACHINE LEARNING MODEL(SOA and ISOA) was presented to exploit multi-modal data comprising of weather and pollution measurements obtained by sensors and picture data gathered by cameras. To deal with the large class imbalance in the data, generative models in combination with typical data augmentation procedures have been utilised to offer a strong prediction capacity to the models. A In order to accomplish the sustainable growth of society and minimise energy consumption, this research offers an improved seagull optimization algorithm (ISOA) paired with an unequal division approach to solve dynamic optimization issues via the analysis of chemical processes.

Finally, the ISOA is paired with an unequal division approach to tackle dynamic optimization issues in order to demonstrate the feasibility and efficacy of enhancing SOA. The viability of the approach is validated by comparing with different algorithms. In this paper,many reference techniques employ fundamentally the uniform discretization of control variables, and the temporal domain has to be separated into several segments. In addition, we partition the temporal domain into fewer segments and apply the unequal division technique of the control variable, which simplifies the computation.

REFERENCES

1. R. Tibshirani, "Regression shrinkage and selection via the lasso", J. Roy. Statistical Soc. Series B (Methodological), vol. 58, no. 1, pp. 267-288, 1996.

2. M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables", J. Roy. Statistical Soc.: Series B (Statistical Methodology), vol. 68, pp. 49-67, 2006.

3. L. Li, X. Zhang, J. Holt, J. Tian and R. Piltner, "Spatiotemporal interpolation methods for air pollution exposure", Proc. Symp. Abstraction Reformulation Approximation, pp. 75-82, 2011.

4. Y. Zheng, F. Liu and H.-P. Hsieh, "U-Air: When Urban air quality inference meets big data", Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 1436-1444, 2013.

5. H.-P. Hsieh, S.-D. Lin and Y. Zheng, "Inferring air quality for station location recommendation based on urban big data", Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 437-446, 2015.

6. M. Dong, D. Yang, Y. Kuang, D. He, S. Erdal and D. Kenski, " PM \$_{2.5}\$ concentration prediction using hidden semi-Markov model-based times series data mining ", Expert Syst. Appl., vol. 36, no. 5, pp. 9046-9055, Jul. 2009.

7. S. Thomas and R. B. Jacko, "Model for forecasting expressway PM2.5 concentration—Application of regression and neural network models", J. Air Waste Manage. Assoc., vol. 57, no. 4, pp. 480-488, 2007.

8. Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, et al., "Forecasting fine-grained air quality based on big data", Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 2267-2276, 2015.

9. X. Zhou, W. Huang, N. Zhang, W. Hu, S. Du, G. Song, et al., "Probabilistic dynamic causal model for temporal data", Proc. Int. Joint Conf. Neural Netw., pp. 1-8, Jul. 2015.

10. K. P. Singh, S. Gupta and P. Rai, "Identifying pollution sources and predicting urban air quality using ensemble learning methods", Atmospheric Environ., vol. 80, pp. 426-437, 2013.

11. C. Rosenberg, M. Hebert and H. Schneiderman, "Semi-supervised self-training of object detection models", Proc. 7th IEEE Workshop Appl. Comput. Vis., pp. 29-36, 2005.