[1.]Shweta K. BORSE

A REVIEW: PREDICTING AIR QUALITY USING DIFFERENT TECHNIQUE

[1.]Dept of Computer Engineering, R. H. Sapat College of Engineering, Management Studied and Research, Savitribai Phule Pune University, Nashik, INDIA

Abstract: Nowadays pollution problems are growing in the world. It has become a big issue because it is harmful to health, it causes health problems like asthma, cancer, heart attack and so on. An air impurity is defined on chemicals present in the environment. If a chemical is higher than the target level of chemicals presents in the air, it is known as an air pollutant. Air pollution levels in most of the urban areas have been a matter of serious concern. In forecasting of pollution, the soft computing techniques are used. Besides, machine learning techniques and data mining techniques namely recurrent neural network, Long Short-Term Memory, gated recurrent unit, artificial neural network, support vector machine, fuzzy logic, genetic algorithm is used in major places worldwide for prediction of the air pollutant.
Keywords: Air Quality Index, Recurrent Neural Network, Long-Short Term Memory, Gated Recurrent Unit, Fuzzy Logic, Artificial Neural Network

1. INTRODUCTION

Some Asian Cities is so bad condition in air pollution, the cities are covered up by fog that slows visibility. According to WHO, the pollutant effect on human health are considered the critical limit, specifically in developing countries like India, China, Pakistan, and Bangladesh. In Asia small cities like Kathmandu, WHO suggested that the PM level be over moderate for some target. The Earliest few years ago, China has been covered by fog with smoggy pollution in their cities, India and Pakistan also include in it. PM is most severely in Delhi. The prediction gives the impurities awareness or the Air Quality Index (AQI). Air pollution predicting container done by join weather prediction with Biochemical transportation system and Environment distributed model. By using a machine-learning algorithm. The prediction takes local production sources (industry or traffic) and isolated sources. The prediction on a historical daily or hourly data and the longitudinal can modify from chunk determination. Forecasts of air quality cover 2 to 5 days Estimating meteorological conditions, there are representations to forecast stages of air quality and air pollution. There are different prediction methods that involve additional difficulty than climate prediction methods. These methods are measured the air impurities concentrate in the airborne. Air pollution arises when the surrounded air containing fumes, dirt, gases in risky to be uncertain to humans and wildlife, plants and resources.

— **Benefits**

Air pollution predicting is a valuable asset on several stages – global, national, community, and individual. Exact predicting services help people proposal ahead, reducing the effect on human health. If public are alert to dissimilarities in the air quality, the effect of impurities on health, a greater probability changes in individual behavior and public strategy, as public need air quality info. Such recognition has the possible to generate a clean atmosphere and also a better resident. Administrations too make usage of initial predicting to initiate procedures to decrease the seriousness of pollution levels.

— **Accuracy in air quality forecasting**

Air smog are powerfully connected with resident climate conditions and surrounded smog emissions. Forecasting air quality, so, not one of the contains the difficulties of climate predicting, it needs information on and data of:

» Resident impurity meditations and emissions.
» Impurity meditations and emissions from different places.
» Movements and probable revolutions of impurities.
» Prevailing air-streams.

The several issues on production in forecasting air quality effect in air pollution predicting existence both particular and objective.

— **Air quality prediction techniques**

There are many such prediction methods, and all need extra complication than climate prediction methods. These methods are measured simulations of how air impurities diffuse in the airborne. The first step to air quality prediction is an outstanding climate prediction. Atmospheric (climate) predicting can be considered into three types: climatology, statistical methods and 3-D models.

≡ **Climatology**

Climatology is based on the guess that the earlier is a decent pointer of the future. This technique is built on the relation among exact climate situations and smog levels, and so on actual one-dimensional.

This technique is frequently stretched to contain the identical of climate patterns to smog patterns. There are several limits to this technique and it is observed as a implement to quantity other predicting systems.

≡ Statistical Methods

The suggestion between climate forms and air quality can be measured using numerical approaches. Some are as follows:

» Classification and Regression Tree (CART):

Classification and regression tree (CART) are proposed to categorize information into different groups. Software classifies variables that associate by ambient effluence level. The information is used to prediction application built on climate situations also on connected impurity awareness.

» Regression Analysis:

Regression analysis evaluations relations among variables. By studying past datasets, organization is completed between smog levels and atmospheric statistical variables. The outcome is an equivalence used for prediction of upcoming effluence levels.

» Artificial Neural Networks:

Artificial neural networks usage adaptive knowledge & pattern recognition methods. Computer established systems are considered to simulate the human brain's capability for pattern recognition. This is uncertainly the maximum suitable technique for predicting smog outstanding toward its multi-dimensional method.

» Three-dimensional (3-D) models

Three-dimensional models exactly signify completely the significant developments that have an effect on outdoor air effluence phase. Three-dimensional method put on the production, transportation, and revolution of air pollution by production usage of some sub method, including:

Emission model:

The longitudinal supply of productions as usual human sources.

Meteorological model:

Produces a trajectory method to forecast the ambient points of effluence with the 3-D atmospheric method.

Chemical model:

Appearances at the revolution of initial pollution into secondary smog to determine the result of the impurity.

Pollution prediction methods are quickly developed and will remain growing in exactness. Exact and available air pollution predictions help increase community perception, allow designed for complex to plan fast, and make available administrations by material for public health warnings [13],[14].

— Air Pollution Accidents

Last few years ago we read or listen to the news about air pollution accidents in the media. We also listen to news about Bhopal Gas Tragedy it should happen cause of four-hour leakage of methyl isocyanate at chemical plant in1984, in this accident 2800 was died in Bhopal, Madhya Pradesh, India.

Table 1. Air Pollution Accidents

Days of fog	Year	Cities	Died People
-	1880	London	-
3	1930	Meuse valley, Belgium	60
9	1931	Manchester, England	592
4	1948	Donora, Penn	20
4	1952	London, England	4000
-	1953	New York	220-240
30	1963	New York	300-405
2	1966	New York	168
3	1991	London	160
12	2017	Punjab, Pakistan	20

LITERATURE REVIEW

Two methods for predicting air quality as namely, lazy learning (LL) and pruned neural network (PNN). PNNs establish a constraint ungenerous method, founded on the deletion of terminated constraints after completely linked ton neural networks; LL is a simple linear forecast system, which does a partial knowledge method each historical data estimate is essential. LL forecaster can be speedily established, then the easiness in the regressors permits equally to the overfitting problems [15].

Air quality estimating with machine learning methods to forecast the hourly of air impurities. Machine learning is capably training a model on large data by using extensive optimization method. They proposed distinguished models to forecast the hourly air smog absorption on the basis of atmospheric data of earlier times by preparing the calculation over 24 hours as a multi-task learning (MTL) [16].

The air quality is poorly affected various methods of smog produced by transport, electrical energy, gas usages etc. The harmful smokes are making a serious for the quality in towns. With growing air smog, effective air quality models collect material data of air impurities and make available valuation of air smog in every zone. From now, air quality estimation and forecast has turn out to be an important research topic. The multi-dimensional features with indeterminate variables, time, location. Authors published examination results air quality estimation with different methods of deep learning, decision trees, artificial intelligence [12].

Spatio-temporal deep learning (STDL) built forecast of the air quality that basically studies multi-dimensional and historical information is establish. A stacked auto encoder method is used to integrate air quality, and it is train on a grasping layer. Associated with time series estimate the models, model can forecast the air quality of places concurrently and demonstrations in all time periods. Support vector regression, auto regression moving average, and spatio-temporal artificial neural network models determines the method of performance air quality forecasts takings a more performance [17].

Air quality prediction is flattering. There are two tools for analysis and prediction, namely numerical and statistical tool. Mathematical and statistical techniques are used for air quality forecast, and other remaining parts used in the physical model. According to using algorithm data will be coded in mathematical form, sometimes they provide the limit for the accuracy so that's why get problem to take accurate accuracy at greatest point (such as taking the cut-off value in prediction) [2]. There are some techniques are used for air pollution prediction. Some areas following:

» Recurrent Neural Network (RNN)
» Long Short- Term Memory Unit (LSTM)
» Gated Recurrent Unit (GRU)
» Artificial Neural Network (ANN)
» Support Vector Machine (SVM)
» Fuzzy Logic (FL).
» Genetic Algorithm (GA).

⎯ Recurrent Neural Network (RNN)

RNN is an open-source machine learning framework that implements Recurrent Neural Network architectures, such as LSTM and GRU. Fabio et al. (2019), proposed to predict the air quality index (AQI) pollution in the Apulia region. Authors have trained different RNN models by using GRU and LSTM neuron cell respectively and forecast the pollution level in some areas by using available data of that particular areas. Using these methods, the outlook on ⎯ air station and weather in the region, it is feasible to alert some days before on the pollution levels.

Estimate approaches based on multiple linear regression, neural network, and recurrent neural network used for the air quality in station are established and associated. The RNN model, forecast the air pollution at station by the prior historical data of the impurities on the model. The forecast models are useful to a actual indoor air quality data-set from tele-monitoring systems data, model shows that the certain variables consume though impact on the forecast presentations. It establishes the RNN that has the capability to model the dynamic system, nonlinear and the forecast outcome of the RNN method shows the good performance and developed interpretability than other forecasting models [18].

⎯ Long Short-Term Memory (LSTM)

The prediction of PM2.5 concentrate using LSTM and RNN. They used TensorFlow tool for LSTM and RNN in python. Environmental Protection Administration of Taiwan data set is used for training (data from 2012 – 2016). Data for the year 2017 is used as test data. Prediction value of PM2.5 is evaluated, and concentrates for the next four hours at some number of stations. Prediction on PM2.5 outcome is poor when PM2.5 does not affect on air pollution [3].

Forecast PM 2.5 smog of air quality station taking the historical data, atmospheric data, climate prediction data. Analytical model be made upon two mechanisms are long short-term memory (LSTM) model the local distinction of PM2.5 pollution and a neural network based three-dimensional combinatory to detention longitudinal dependence between the PM 2.5 pollution of station areas and that of neighbor stations areas. Estimate model on a dataset records of air quality stations in Beijing and associate it with artificial neural network and LSTM on the similar dataset. The outcomes show that LSTM-FC neural network gives a best performance [19].

⎯ Gated Recurrent Unit (GRU)

GRU is simple than LSTM. GRU contain two types of gates as Update Gate and Reset Gate. Three different GRU in RNN to retain the form and reduce the parameter in reset and update gate. The GRU model on IMDB and MNIST data set. Gates occur to the recurrent state al about the other signals. They use GRU1, GRU2, and GRU3 RNN. Test accuracy is similar to three sequence length of them [4]. Beijing weather data and cluster data used as input vector. The training data is divided into four sections according to the season of winter, autumn, summer, and spring. Four models effect on forecasting PM 2.5 analysis by test data sets. After taking demonstration and persistent adaption of parameter, the forecasting errors and accuracy of models are examined and compared, then its work ability and advantages are verified by this method. Forecasting accuracy is built on the high GRU model, this model is effective for timeseries prediction of air pollution [5].

⎯ Artificial Neural Network (ANN)

One-year data used for training and two-week data separately used for testing. Their huge amount of data removed from the training data set, compare with time-lagged model [7].

ANN is a data analysis tool used for optimizing the air quality data. It takes the three sets of data as test, train, and validation of models. Data set divided randomly training data set takes 75%, Testing takes 15%, and Validation data set takes 15% automatically. The result of air pollution in the atmosphere using different forecasting [6].

⎯ Support Vector Machine (SVM)

SVM is a supervised learning, it used for regression or classification challenges.

Supervised learning (labeled data) used as training data. Air quality is divided into two types using Support vector machine as namely good or harmful. Mathematical formula Cumulative Index (CI) basis on four pollution as namely SO2, PM 2.5, PM 10, NO2. Possibility of input for SVM classifier value tested on real data. Polynomial kernel used in SVM that gives great accuracy and good performance more than forecasting. In the tuning model, accuracy is used primary metric [8].

─ Fuzzy Logic (FL)

To predict air quality in urban areas and other cities by integrate CEP (Complex Event Processing) technology and fuzzy logic. Pollution data derive d from weather prediction and CEP engine by territory specialist. New fuzzy rule and input vector easily include in the FIS model and also modified a plan to fewer air pollution [9]. Develop electrostatic filter form on fuzzy logic to trim air pollution of PM. Electrostatic filter is O3 generate consist of plate corona expel and high volt DC generator. In experimental result s, they get trim Particulate matter (PM) of PM 10 in ten minutes by 55%, 68%, and 75% respectively for one, two, and three ozone generators [10].

─ Genetic Algorithm (GA)

GA for selecting the starting weight and also used to learning something. It estimates in nature and then search for the final solution. GA calculates the starting weight. By using GA optimize the ANN nature. GA and ANN give the best performance by using a hybrid method [11].

Weather Research and prediction using California PUFF model. GA is used for the design of the best examine network [12].

SYSTEM ARCHITECTURE

Recurrent neural networks are one of the simplest models using the task of forecasting the afterward is a categorization based on the previous ones. Feedback is memorized portions of the involvements and used to make accurate predictions. Training a typical neural network involves the following steps:

\# Input an air quality and weather from a dataset.

\# The network will take that air quality, and weather, and apply some complex computations to it using randomly initialized variables (called weights and biases).

\# A predicted result will be produced.

\# Comparing that result to the expected value will give us an error.

\# Propagating the error back through the same path will adjust the variables.

\# Steps 1–5 are repeated until we are confident to say that our variables are well-defined.

\# A predication is made by applying these variables to a new unseen input.

The figure 1 shows the workflow of implementation to train the RNN. RNN can be trained to estimate CAQI level using air and weather data from different stations. Data split into two parts as the RNN model is used for trained and second data is used as RNN test data. In the training process, it takes all available data from datasets, then consider as test arbitrary sequential days. In the second test request to RNN, to make "blind forecast". First, randomly choose an air station, and then train the RNN to use data in a specific range. Estimate the direction between the selected station and other, and analyze weather state, trained to forecast CAQI level. Information about weather state of before days, depends on wind direction, RNN tries to evaluate the value of every impurity and CAQI [1].
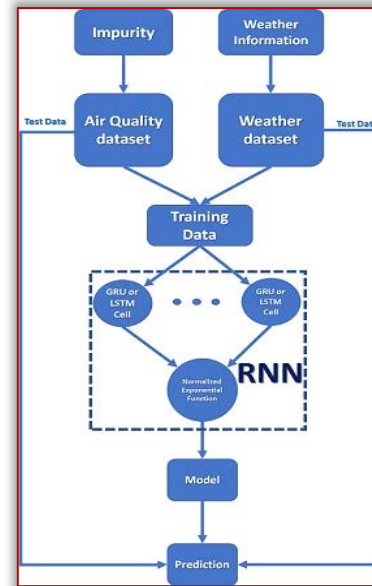

Figure 1. RNN model with Training and Test Data

CONCLUSION

In this paper, we study of air pollution and air quality forecasting using different algorithms. There is indisputable evidence of the unfavorable impacts of low air quality on the environment also on human health. Exact estimating benefits to public strategy gaining, reduce the belongings on fitness and the outlays are linked. If community are alert for differences in the value of the airborne, they blow, the result of impurities on fitness as well as meditations probable to basis opposing effects and activities to restrict smog. In forecasting of pollution, soft-computing methods are used. We also see that machine learning techniques and data mining techniques namely long short-term memory, recurrent neural network, gated recurrent unit, artificial neural network, support vector machine, fuzzy logic, genetic algorithm is used in major places worldwide for prediction of the air pollutant.

References

[1] Cassano. F., Casale, A., Regina, P., Spadafina, L., & Sekulic, P. (2019). A Recurrent Neural Network Approach to Improve the Air Quality Index Prediction. Ambient Intelligence- Software and Applications –,10th International Symposium on Ambient Intelligence, 36–44

[2] Niharika, Venkatadri M*, Padma S. Rao / (IJCSIT), A survey on Air Quality forecasting Techniques, International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 103-107.

[3] Tsai Y.-T., Zeng, Y.-R., & Chang, Y.-S., Air Pollution Forecasting Using RNN with LSTM. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)

[4] R. Dey and F. M. Salemt, "Gate-variants of gated recurrent unit (GRU) neural networks," in Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS), Aug. 2017, pp. 1597–1600.

[5] Xinxing Zhou, Jianjun Xu, Ping Zeng, and Xiankai Meng, Air Pollutant Concentration Prediction Based on GRU Method, Laboratory of Software Engineering for Complex Systems, College of Computer, National University of Defense Technology, Changsha, China, IOP Conf. Series: Journal of Physics: Conf. Series1168 (2019) 032058

[6] M Venkatadri, A Novel Air Quality Prediction Model Using Artificial Neural Networks, International Journal of Engineering Research, 2014 - researchgate.net

[7] Madhavi Anushka Elangasinghe, Naresh Singhal, KimN. Dirks, Jennifer A. Salmond, Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis, Atmospheric Pollution Research (2014) 696-708.

[8] Saxena, A., and Shekhawat S. (2017). Ambient Air Quality Classification by Grey Wolf Optimizer Based Support Vector Machine. Journal of Environmental and Public Health, 2017, 1–12

[9] Macia. H., Díaz, G., Boubeta-Puig, J., Valero, E., & Valero, V. (2019). Combining Fuzzy Logic and CEP Technology to Improve Air Quality in Cities. Computational Science – ICCS 2019, 559–565

[10] Dewi Nurhaji Meivita, Muhammad Rivai, Astria Nur Irfansyah, Development of an Electrostatic Air Filtration System Using Fuzzy Logic Control, Vol.8 (2018) No.4ISSN: 2088-5334.

[11] Hao, Y., Xie, S., Optimal redistribution of an urban air quality monitoring network using atmospheric dispersion model and genetic algorithm, Atmospheric Environment (2018)

[12] Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie, Air Quality Prediction: Big Data and Machine Learning Approaches, International Journal of Environmental Science and Development, Vol. 9, No. 1, 2018

[13] "National Air Quality Monitoring Program". CPCB. Archived from the original on 2 May 2013. Retrieved 31 May 2013.

[14] Climate change in Asia http://recap.asia/climate-asia/Air-Pollution-in-Asia.html

[15] Corani, G. (2005). Air quality prediction in Milan: feed-forward neural-networks, pruned neural networks and lazy learning. Ecological Modelling, 185(2-4), 513–529

[16] Zhu, D., Cai, C., Yang, T., & Zhou, X. (2018). A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization. Big Data and Cognitive Computing, 2(1), 5

[17] Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. Environmental Science and Pollution Research, 23(22), 22408–22417

[18] Kim, M. H., Kim, Y. S., Lim, J., Kim, J. T., Sung, S. W., & Yoo, C. (2010). Data driven prediction model of indoor air quality in an underground space. Korean Journal of Chemical Engineering, 27(6), 1675–1680

[19] Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere, 220, 486–492