International Journal of Scientific & Engineering Research, Volume 10, Issue 12, December-2019
ISSN 2229-5518

Air Pollution Monitoring Using Artificial Neural Network (ANN)

Ruchi Bhavsar

Abstract—— Air Pollution and its precaution has been a scientific challenge since past few decades. And it still remains an endless global problem. Due to the growing population and increase in settlements in urban areas, the amount of pollutants in air increase day by day. Affecting human's respiratory and cardiovascular system, they are cause for increased mortality and increased risk for diseases for the population. Many efforts have been taken by the governmental bodies to understand and predict the AQI (Air Quality Index) to further improve public health. The most important step in prediction, no doubt, is to develop a predictive model of air quality status, which will help in management of the environment and also to create a sense of awareness among people. Air quality prediction is the key factor for monitoring air pollution.

Index Terms— Air pollution, AQI, Air quality index, NO2, Particulate matter, SO2, Multilayer Perceptron.

1 Introduction

Artificial Intelligence and machine learning has seen a sudden rise in the past few years. In the field of artificial intelligence where the machine makes all the decisions on its own rather than having to program the machine for every specific need of the user, has started integrating itself in every aspect of our life. Artificial intelligence and machine learning has become the key focus of every company and organization, from start-ups to well known vendors. The procedure of machine learning is that the machine first senses the environment and the factors related to it. It then gathers information and makes appropriate decisions according to it. One of the main reasons why machine learning was chosen to predict air quality index and other harmful pollutants was this ability of adapting of machine learning (ML) algorithms.

The air is filled with pollutants and they can be in the form of liquid, gaseous or solid. These pollutants have a varying concentration, which could determine its toxicity. If the concentration of some pollutants is high, it could severely affect human beings, plants, animals and environment. There are multiple sources for excess cause of air pollutants such as indoor activities as well as outdoor activities. Many domestic activities like cooking, burning of bio-fuels and heating produce indoor air pollutants. There are two types of outdoor air pollutants namely, primary and secondary air pollutants. The pollutants that are released by industrial activities and fuel combustion into atmosphere are primary pollutants, whereas the secondary pollutants are a result of reaction in atmosphere. The various harmful air pollutants are Oxides of nitrogen, particulate matter, carbon monoxide and carbon dioxide.

- *Ruchi Bhavsar is currently pursuing bachelor's degree program in Information Technology in Thakur College of Engineering and Technology, Mumbai, India. E-mail: ruchibhavsar98@gmail.com*

2 Background

The detailed investigation with a plan to locate a straightforward non-direct model for a genuine neuron prompted the advancement of Artificial Neural Network (ANN) hypothesis by McCulloh and Pitts in 1943. McCulloh and Pitts attempted to show the bio frameworks utilizing nets of basic legitimate tasks.

This advancement brought an extraordinary enthusiasm from different analysts and researchers everywhere throughout the world. In this way a few ANN-based models in various fields were found. The innovation however had lost its force in the late 1969 till 1986 when the back-proliferation of blunder was found. ANN-based models have been effectively executed in various disciplines extending from: Medical, car, safeguard, hardware, aviation, stimulation, monetary, etc.

In view of ANN approaches, Crawford, 2000 detailed the intense a ruptured appendix investigation. ANN is a vital part of smart based frameworks, planned unmistakably to improve the presentation of customary figuring methods. The greatest disadvantage related with the alleged customary strategies is the failure to learn and recognize designs in unique frameworks. Accordingly the need to wipe out this inadequacy through learning is demonstrated fundamental.

At the end of the day, however the factual strategies do give sensible outcomes, these are basically unequipped for catching unpredictability and non-linearity of contamination climate connections. To defeat this negative mark of measurable strategies, Artificial Neural Network (ANN) system, the third approach created lately has become the focal point of a lot of consideration, to a great extent since they can deal with the non-linearity and have been utilised to show position fixations with promising outcomes.

International Journal of Scientific & Engineering Research, Volume 10, Issue 12, December-2019
ISSN 2229-5518

3 Related Work

Kaminski et al. in 2008 analyzed that an effective instrument for air quality in urban cities is neural network as it predicts with low errors. In the ideally built models, prediction errors were just 1.9% in testing and 1.4% in training.

Barbes et. al. in 2009 proposed a model for concentration prediction of inorganic airborne pollutants: H2S-SO2, NO-NO2-NOx, CO (CO2) and PM10 (particle matter with an aerodynamic diameter of 10ìm or less) from a risk area (two industrial area - IA) and an urban area (UA) from Constanta. The examination between the outcomes from genuine estimated information from urban region and the results of simulated values gave a small error of 0.42. This further depicts the efficiency and validity of the given method in the evaluation of various pollutants.

Chabaa et al. (2010) analyzed the internet traffic data over IP networks by development of an artificial neural network (ANN) model based on the multilayer perceptron (MLP). Developed models, using the LM and the Rp algorithms, can success-fully be used for analyzing internet traffic over IP networks, and can be applied as an excellent and fundamental tool for the management of the internet traffic at different times.

Afzali et al. (2012) studied on the potential of Artificial Neural Network Technique in Daily and Monthly Ambient Air Temperature Prediction. The mean, minimum and maximum ambient air temperature during the years 1961-2004 was used as the input parameter in Feed Forward Network and Elman Network. The values of R, MSE and MAE variables in both networks showed that ANN approach is a desirable model in ambient air temperature prediction, while the mean temperature of the next day and maximum of the next one month are more precise using Elman network.

Mahmoudzadeh et al. (2012) described a comprehensive computer modelling based on the current and previous related information for further study, analyses and decision making is of paramount importance. For both training and testing phase, mean square error is used for performance evaluation. Comparison between the functionality of hybrid ICANN and the mentioned MLP network provides the fact that the ICA-NN was superior in terms of reliable performance with acceptable accuracy for CO pollutant prediction.

Azid et al in 2013 came to a conclusion that the management of atmosphere will be done efficiently when ANN is used for problem solving. He took into consideration the Peninsular Malaysia which is based on the principal component analysis and ANN for the prediction of air quality index (AQI).

4 Proposed Work

Step 1: Collect or get the data and look for parameters like sulphur oxides, nitrogen oxides, particulate matter, etc. There are two types of particulate matter that are particularly harmful for humans, they are-PM10 and PM2.5. Using these parameters, we will be able to predict the pollution.

Step 2: The existing data parameters and their air quality parameters are fed to the training set. The resulting output of the training set is the Air Quality index or AQI. Data has to be provided in time series in the form, day wise.

Step 3: Fetch for historical data and then train this data using the Multilayer Perceptron (MLP). MLP is a neural network model that plots set of data inputs into a set of appropriate outputs. It also provides us with a model.

Step 4: The trained model will help in providing new predicted values of Air Quality Index for the next day. Furthermore, the air quality predictions for next month or even a year can be obtained using the MLP model. The future data points will also be used for recognition using the MLP model.

Step 5: The predicted values of the parameters can be displayed city-wise. We store all historical data into the database at the time prediction of the AQI. And using this data, city-wise analysis graphs will be provided.

Step 6: Certain factors that add to the increasing amount of pollutants, like pollution from vehicles, incidents that caused excess pollution, are taken into consideration area wise. Using these will further help to precisely determine the Air Quality Index (AQI).

Step 7: Area wise, state wise or even country wise prediction could be made. This will help in creating awareness among people and also take precautions.

The Air quality index obtained will be governed by some value. This value can help determine if the air is fit for a living being or not. An already generated table showing the range of values that will decide the air quality is given below.

International Journal of Scientific & Engineering Research, Volume 10, Issue 12, December-2019
ISSN 2229-5518

TABLE 1
AIR QUALITY INDEX RATINGS

Index	Rating	Comment
0-50	Good	No risk
51-100	Moderate	Can be dealt with
101-150	Unhealthy for sensitive people	Children and elderly are affected
151-200	Unhealthy	Not fit for anyone
201-300	Very Unhealthy	Serious health problems
301-500	Hazardous	It is very dangerous

5 DATA PRE-PROCESSING

Data preprocessing is the arguably one of the most important section of the project towards building an effective and accurate model. There are many steps in data pre-processing that need to be completed for a desired outcome. The various steps are:

i) Data Collection

ii) Data Cleaning

iii) Feature Scaling

5.1 Data Collection

Firstly, we retrieved already formulated data about various pollutants such as No2, So2, Particulate matter (PM2.5 and PM10) and carbon monoxide. The values of these parameters were given region wise and were expressed in µg/m3 (micrograms per cubic meter). Pollutant data is stored in the dataset for easy access and manipulation. Data will be used as a training set for complete learning of the machine. As the data taken for training from the website was not clearly defined and labeled, there was a need for cleaning it.

5.2 Data Cleaning

An unclean data affects the accuracy of the model. The main purpose of data cleaning is to remove errors and inconsistent data in the training set. This will enable us to have a better view of the data and to help in developing a reliable model. The first step towards data cleaning and preprocessing is to import libraries. Library is a tool that when given an input gives a desired output. The three libraries out of many, in Python, that are extremely important are Numpy, Pandas and Matplotlib.

- Numpy- Numpy library that deals with all the mathematical equations in your project. It can be imported by using 'import numpy as np' in the code.
- Pandas- Pandas library is the best for importing and managing datasets. It can be imported by using import pandas as pd' in the code.
- Matplotlib- This library is useful when making charts. It can be imported by using 'import matplotlib.pyplot as plt' in the code.

Out of these three, Numpy and Pandas are the best options for data preprocessing. The data is split into two types:

i) Training set

ii) Testing set

The values in the training data set are used to train the model and learn how to map inputs to outputs from examples in the set. The testing data is the used to test whether the model performs efficiently with the provided value and give appropriate output. These values of training and testing sets cannot be interchanged to avoid overfitting problem.

5.3 Feature Scaling

In real world applications, feature scaling has become one of the most important step in data-processing. When using a three layer feed forward neural network, removal of unwanted data is necessary. Most of the times the data that you are presented with has a very high range meaning the difference between the minimum value and maximum value in the dataset is higher for one single parameter. These differences in scales across inputs may increase the difficulty. It often causes the model to be unstable, will suffer from poor performance and result in higher generalisation error. To reduce this difference and make the values much closer to each other, feature scaling is used. It can be achieved by using Normalisation techniques. These normalisation techniques can also help in improving the performance of Multilayer Perceptron Model (MLP).

Normalisation

Normalisation is a rescaling technique where the all the values are between 0 and 1. There are two types of normalisation:

i) Z-score

ii) Min-Max

International Journal of Scientific & Engineering Research, Volume 10, Issue 12, December-2019
ISSN 2229-5518

We will be using Min-Max technique as it is considered to be a good practice in comparison with z-score.

6 Algorithms Used

Mainly two algorithms are used in this project and they are:

i) Linear Regression
ii) Multilayer Perceptron (MLP)

6.1 Linear Regression

A straight or linear positive or negative line that justify the observational points between the dependent and independent variable is known as linear regression. The manipulation and changes are made to the independent variable and accordingly the dependent variable is altered. When you have one independent variable, it is known as simple linear regression. And when you have more than 1 independent variable, it is known as multiple linear regression. The equation representing multiple linear regression is:

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon$$

Where, y is the dependent variable that is tone estimated, X are the independent variables and ε is the error term. β (1,2,..,k) are regression coefficients.
The value of β (regression coefficient) can be found out by:

$$\hat{\beta} = (X'X)^{-1}X'y$$

6.2 Multilayer Perceptron (MLP)

Multilayer Perceptron, shortly known as MLP, is a part of Artificial Neural Network where parameters are fed as input and resulting output is given. It constitutes a group of interconnected processing units each associated with a learning rule. This unit is known as neuron. Its mathematical representation is:

$$S = f\left(\sum_{j=1}^{n} e_j w_j\right)$$

Where ej is the j-th input to the neuron, wj is the weight associated with ej, n is the number of inputs, f is the activation function (tanh or sigmoid) and s is output of the neuron.

Weights

Weights are coefficients on the input. Similar to linear regression, each neuron has a bias which can thought of as an input that has the value 1.0 and need to be weighted. Weights are initialised to small random values at first usually between 0-0.3 for less complexity.

Activation Function

Activation function is also known as transfer function. The weighted inputs are summed and passed through the activation function. It is given this name for its property of governing the threshold at which the neuron is activated. Usually non-linear activation functions are used such as the sigmoid function which has values between 0 and 1.

MLP has three layers in a basic network and they are input layer, hidden layer and the output layer. There is only one input and output layer but the network can have multiple hidden layers. This algorithm contains numerous coating of nodes, every layer in the network is connected to the layer before it for its output with an exception for input nodes.

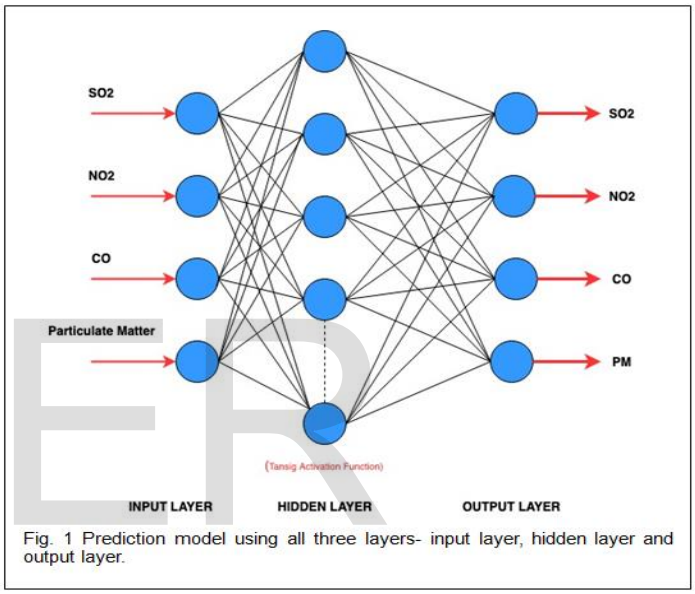

Fig. 1 Prediction model using all three layers- input layer, hidden layer and output layer.

A feed forward architecture of MLP network is used in this project. A feed forward network is a network where the information flow is only one way and no feedback is provided. The parameters of air pollution such as NO2, SO2 and particulate matter are fed to the input layer of network. Computations are performed in the hidden layers and the final output is given by the output layer.

7 Conclusion

The work done in this project gives an overview of how the model, for monitoring air quality, can be structured and built. The various algorithms used will be helpful in creating an effective model for prediction pollutants. Thus, it is clearly observed that using Artificial Neural Network is the most appropriate method to be used for prediction. With progressively serious air contamination, it is imperative to foresee air quality precisely for giving appropriate activities and controlling systems so the unfriendly impacts can be limited. Using historical data has helped in fruitful training and testing of the model. It can further be improved by the use of additional variables such as the effects of pollution caused by multiple

International Journal of Scientific & Engineering Research, Volume 10, Issue 12, December-2019
ISSN 2229-5518

industries on the values of SO2, NO2 and particulate matter, ambient temperature, wind speed and relative humidity. Or there could be a different reason altogether such as the burning of forests which would have a large effect on the prediction. Lastly, the accuracy of this model can be further increased.

REFERENCES

[1] Afzali M., Afzali A., Zahedi G.M., "The Potential of Artificial Neural Network Technique in Daily and Monthly Ambient Air Temperature Prediction," International Journal of Environmental and Science Development. 3(1):33-38.)

[2] Azid A., Juahir, H., Latif, T.M., Zain, M.S., Osman, R.M, "Feed-Forward Artificial Neural Network Model for Air Pollutant Index Prediction in the Southern Region of Peninsular Malaysia". Energy Journal of Environmental Protection. 7(4): 1–10. 2013.

[3] Barbes, L., Neagu, C., Melnic, L., Iliec, Velicum, "The Use of Artificial Neural Network (ANN) for Prediction of Some Airborne Pollutants Concentration in Urban Areas". Journal of Rev. Chem. 60 (3): 301–307. 2009

[4] Chabaa, S., Zeroual, A., Antari, J., " Identification and Prediction of Internet Traffic Using Artificial Neural Networks" Journal of Intelligent Learning Systems and Applications. 2 : 147–155. 2010.

[5] Kaminski, W., Skrzypski, J. and Szakiel, J.E., "Application of Artificial Neural Networks (ANNs) to Predict Air Quality Classes in Big Cities" 19th International Conference on Systems Engineering. 19: 135-140. 2008.

[6] Mahmoudzadeh, S., Othma, Z., Yazdani, M.A. and Bakar,"A.ACarbon Monoxide Prediction Using Artificial Neural Network and Imperialist Competitive Algorithm" Journal of Basic and Applied Sciences. 7(4): 735–744. 2012.