AQI: PREDICTION AND OPTIMIZATION TECHNIQUES

1 Sushma Kumari, 2 Ifshita Chaudhry, 3 Shruti Sharma, 4 Preeti Sethi

1 Student (CE), J. C. Bose University of Science and Technology, YMCA Faridabad
2 Student (CE), J. C. Bose University of Science and Technology, YMCA Faridabad
3 Assistant professor (CSE), J. C. Bose University of Science and Technology, YMCA Faridabad,
4 Assistant professor (CSE), J. C. Bose University of Science and Technology, YMCA Faridabad.

Abstract: This paper covers all optimization techniques that can be used for better classification for prediction of air quality. Further, the comparative analysis is made among the techniques and the best one is chosen on the basis of its accuracy rate. There are various measures that have been reviewed in this paper which had been used for the prediction of air pollution pollutants. Some of them are Deep Learning, Machine Learning, Feed Forward Neural Network, and so on.

Keywords: Air Quality Index (AQI), Genetic Algorithm, Optimization Techniques.

1. INTRODUCTION

Nowadays, Air Quality Prediction is necessary because of many reasons. Let us first understand what Air Quality Index is. An **air quality index** (**AQI**) is used by government agencies to communicate to the public how polluted the air currently is or how polluted it is forecasted to become. Public health risks increase as the AQI rises. Different countries have their own air quality indices, corresponding to different national air quality standards. Since the beginning of industrial revolution and more rapid acceleration from the past several years as more countries have embarked on rapid development, the composition of the atmosphere has been gradually changing and thus this variation in the composition in the constituents of the atmosphere results in air pollution. Thus, Air Pollution can be defined as the presence of substances in atmosphere in insufficient concentrations so that they are said to be a threat to almost all living beings as well as to the environment and the climatic conditions. In other words, air pollution destroys the comfortable enjoyment of life, property and environment.

1.1 Predictive Analytics process:

a. Define project: Define the project that are going to be used .

b. Data collection: The data set can be collected from various sources such as a file, database, sensor and many other such sources but the collected data cannot be used directly for performing the analysis process 33 as there might be a lot of missing data, extremely large values, unorganized text data or noisy data. Therefore, to solve this problem Data Preparation is done.

c. Data analysis: Data analysis involves transforming raw data into an understandable format. Real world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data analysis is a proven method of resolving such issues. Following steps are performed:

1. Wrangle data and prepare it for training.
2. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)
3. Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data.
4. Visualize data to help detect relevant relationships between variables or class imbalances or perform other exploratory analysis. 5. Split into training and evaluation sets.

d. Statistics: Statistical analysis enables statistical approach towards the model.

e. Modeling: Different algorithms are there such as ANN, MLP, CNN, LSTM, CNN-LSTM, Encoder decoder, and Convolution LSTM. We have monitored and evaluated about the algorithms in the above chapter. The best algorithm will be selected for further steps.

f.Deployment: Deployment mode**l** represents the exact category of cloud environment based on proprietorship, size, and access and also describes the nature and purpose of the cloud. Most organizations implement cloud infrastructure to minimize capital expenditure & regulate operating costs.

g. Model Monitoring:independence, Speed and Accuracy. Monitoring is an important element of managing credit risk as it allows users, management, regulators and other stakeholders to be confident in relying on the model and its ratings to manage risk.

1.2 Optimization:

Optimization is finding an alternate with most cost effective or highest achievable performance under the given constraints, by maximizing desired factors and minimizing undesired ones.

Let us assume that a data scientist has an image dataset through an image classifier is to be created. Firstly, the dataset is to be normalized. After that, we have to find K which is the K-Nearest Neighbor (KNN) and is an important parameter.

Assume that initial value of k is 3 and learning process is started by scientist. After that the model generation reached a classification accuracy with 85%. Now, the question arises whether that percentage is acceptable or not for better classification. Explicitly, we cannot say that 85% is the best accuracy rate until we conduct some other experiments. We need to change the value of k in KNN algorithm and notice where the classification accuracy varies and how we can find the best value of k that maximizes the classification performance. This is nothing but optimization.

In optimization, some values are initialized for the variables which will be used further in analysis. As these values may not be the best one to use, they should be changed until we get the best one. In some cases, these values are created by using mathematical

functions that cannot be solved manually. As a result, there are different optimization techniques suggested in different operation research (OR) to do such work of optimization.

2. LITERATURE REVIEW

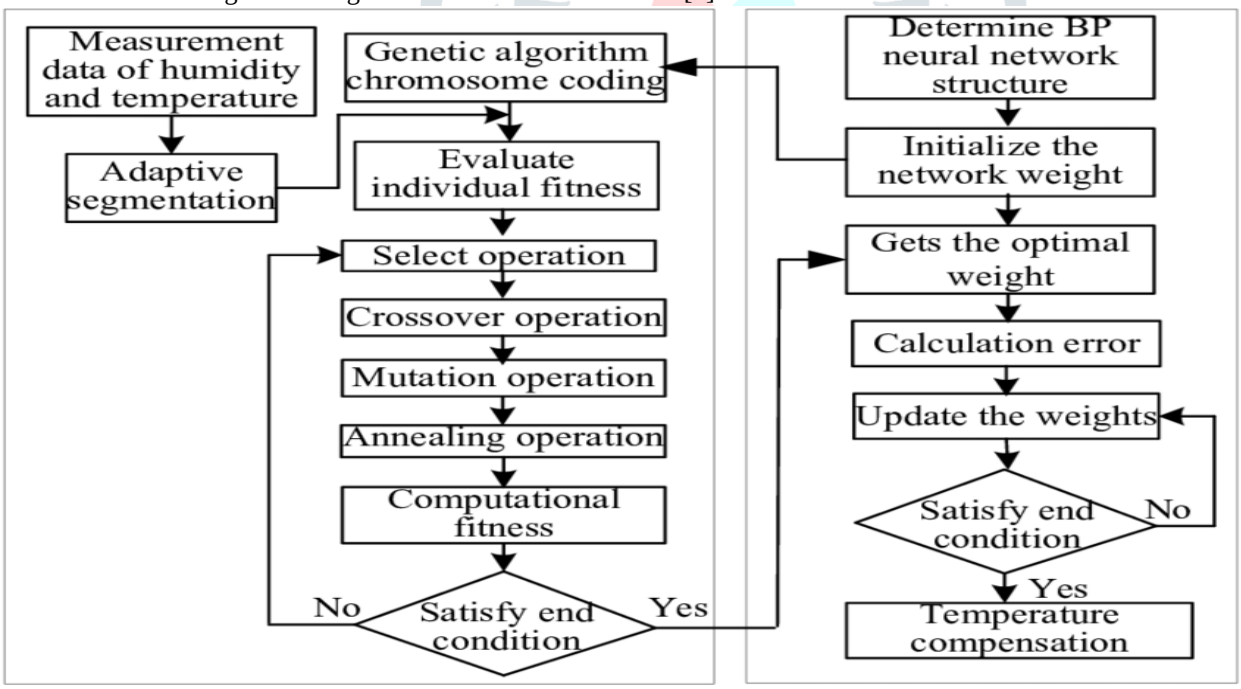
The prediction of air quality is becoming essential for minimizing the environmental imbalances which further effectively addresses the air pollution. There are variety of tools such as numerical and statistical are available for prediction as well as for the analysis of pollution. The drift of advanced computing/analysis techniques from traditional computing methods to recent soft computing techniques effectively addresses the air quality prediction. The traditional approach for air quality prediction uses mathematical as well as statistical techniques. In these techniques, initially a physical model was designed and then the data is coded with some differential mathematical equations.

But such methods do not provide limited accuracy because they were unable to predict the extreme points i.e. the pollution maximum and minimum cut-offs cannot be determined. Also, such methods were lengthy and are inefficient. Hence with the advancement in technology and research, a methods have been proposed i.e. Artificial Intelligence (AI) techniques can be used for prediction purposes. Among various types of soft computing techniques, following are the major air pollution predictive model techniques below:

- Genetic Algorithm (GA).
- Artificial Neural Networks (ANN).
- Support Vector Machines (SVM).
- Fuzzy Logic (FL).
- Hidden Markov Model (HMM).
- Particle Swarm Intelligence (PSI).
- Hybrid soft computing techniques (HSC).

Genetic Algorithm is based on Darwin's Theory [1]. It begins with an arbitrary created individual population and then the fitness is evaluated and childrens are selected from the individuals. Genetic algorithm effectively addresses the change of the accumulation of the surrounding atmosphere and predict the thickness of the air pollutants [2] and are also applied to extract the optimal feature subset of a large database containing pollutant concentration measurements, which feeds to a nearest neighbor algorithm in order to predict the daily maximum concentration for pollutants [4] .In another paper they are used to optimize the accuracy rate of AQI prediction on Backpropagation neural networks [3].

Genetic Algorithm generally used to speed up the search problems by relying on bio-inspired operators such as mutation, crossover and selection process. It will help us to obtain better performance and accuracy of result. The basic objective of Genetic Algorithm is to find the initial weight for using the Artificial Neural network [5].



In paper [8] machine learning approach has been used for prediction of air pollutants such as ozone, particulate matter and sulfur dioxide. The prediction has been made for one hour. They proposed a model which can predict hourly concentration of pollutants using a multi task learning (MTL) problem and also used various normalization techniques which helped to achieve better performance and accuracy rate.

Masoume Asghari Esfandani and Hossein Nematzadeh [9] used ANN-BP (Artificial Neural Network with Back Propagation) hybrid GA(Genetic Algorithm) and BP-PSO (Back Propagation with Particle Swarm Optimization) to evaluate the predictions and concluded that the BP-PSO algorithm provide higher accuracy rate as compared to others.

With the pioneering work of McCulloch & Pitts, Artificial Neural Networks (ANN) has its roots in wide interdisciplinary history from the early 1940's [5]. ANN raised as a mechanism to mimic the human's brain processes. It is an intelligent system which has the capacity to learn, memorize and create relationships among the data. ANN is made up by the simple processing units, the neurons,

which are connected in a network by a large number of weighted links where the acquired knowledge is stored and over which signals or information can pass[1]. The prediction of air quality, effectively addressed by the prediction of various air pollutants like Sulphur, Nitrogen, Carbon Monoxide, Ozone, Suspended Particulate Matter (SPM) divide the data set into training , validation and verification [3]. ANN effectively addresses the prediction of Sulphur Dioxide distribution and the future concentration in the air by modeling the Sulphur Dioxide concentration and its distribution from the air pollution station [2].

As we increase the time of prediction the rate of negative accuracy rate have been seen due to varying amounts of temperature, dust, humidity, noise which are the key factor for the air pollution prediction. Masoume Asghari Esfandani and Hossein Nematzadeh [6] used ANN-BP (Artificial Neural Network with Back Propagation) hybrid GA(Genetic Algorithm) and BP-PSO (Back Propagation with Particle Swarm Optimization) to evaluate the predictions and concluded that the BP-PSO algorithm provide higher accuracy rate as compared to others.According to the Chao Zhang, Junchi Yan, Yunting Li, Feng Sun, Jinghai Yan, Dawei Zhang, Xiaoguang Rui, Rongfang Bie machine learning [7] can be used to predict the air pollutants particulates which helps to define the multi-channel ensemble learning frameworks which uses the Deep Boltzman Machine as the basic building block of learning system. They used the technique based on data of Beijing China and also used in the web service system as well.

The main findings from the above literature are mentioned below in the table 1.

Table 1: Summary of Literature Review

S.No.	Name of Author	Methodology	Results	Pros	Cons
1.	Wang Zhenghua et al. [6]	Back propagation, Genetic algorithm (GA)	1. Accuracy rate is improved. 2.The method is complicated and computations are expensive	1.The accuracy rate of AQI achieved is 80.44% and level is 82.5%. 2.Author propose AQI prediction model of BP neural network based on genetic simulated annealing algorithm optimization	Single Neural Network is slow convergence.
2.	Bin Mu et al. [3]	Principal Component, Genetic Algorithm (GA), Back propagation	More input data are required in network in order to have better accuracy and can use swarm intelligence algorithms for better accuracy.	BP-PCO found to be a data with more accurate data with that of short term predictions as well.	Cost of setting the system is high.
3.	Niharika et al.[]	Genetic Algorithm, Artificial neural network, fuzzy logic, soft computing	In this paper a study was carried out on various Air quality prediction techniques. Here soft computing plays vital role.	Predicts main causes of pollution and factors responsible to minimize it	Complex to design .
4.	Surajit Chattopadhyay et al. [10]	Author propose the prediction of mean monthly total ozone concentrations over the calculated period	Need to predict the value of all months in order to find the best result of the algorithms	Here single hidden layer and Two hidden layer are considered with Back propagation, where Two hidden layer perceptron founds to	Predicting all months value is difficult task

				be best for the prediction of Ozone.	
5.	Ebrahim Sahafizadeh et al.[11]	Data mining technique to Predict dusty days.	Author extracted the data of 53 years and generated decision rule for the given data by using the air pressure, humidity and dusty days of January, February and March of each year	Using Clementine software and author classified dusty days into 5 classes and then exported decision rules to predict each year classification	Limited to one city.

3. CONCLUSION

With the main objectives of this paper is to undergo study of various techniques used in prediction of air pollutant particles and optimize it with best PCA and genetic algorithm so that we would be able to increase the accuracy level of the prediction values to obtain best result which would help our Government system to take preventive measures to avoid this pollution. Since it is found that the longer the time of prediction the lesser the accuracy rate of result, thus with the many techniques that have been undergone concludes to use shortest time span of time so that we would obtain an adequate result with PCA and Genetic optimizations to obtain best and precise result.

4. REFERENCES

[1] S. ShrutiTikhe, Dr. Mrs. K.C. Khare, Dr. K.N. Londhe,2013, "Forecasting Criteria Air Pollutants Using Data Driven Approaches; An Indian Case Study. " IOSR Journal Of Environmental Science, Toxicology and Food Technology, Vol. 3,PP.01-08, ISSN:2319-2399

[2] Shine –Wei Lin, Chih –Hong Sun, and Chin-Han Chen , "Temporal data mining using Genetic Algorithm and Neural Network – A case study of Air pollutant forecasts".

[3] Bin Mu, Site Li, Shijin Yuan, 2016, "Air Quality Forecast Based On Principal Component Analysis - Genetic Algorithm, Back Propagation Model", International Journal of Environmental and Ecological Engineering, vol. 10, no. 08.

[4] Elias Kalapanidas and Nikos Avouris, Feature Selection using a Genetic Algorithm applied on an Air Quality Forecasting Problem. Available www.hci.ece.upatras.gr/pubs.../C54_Kalapanidas_Avouris_BESAI_ 2002.pdf

[5] N.Sharma, K.K. Chaudhary, CV ChalapatiRao, 2005, "Vehicular pollution modeling using artificial neural network technique - A review", Journal of Scientific and Industrial Research, Vol 64, pp. 637 647.

[6] Wang Zhenghua, Tian Zhihui, 2017, "Prediction of air quality index based on improved neural network", International Conference on Computer Systems, Electronics and Control (ICCSEC).

[7]Chao Zhang, Junchi Yan, Yunting Li, Feng Sun, Jinghai Yan, Dawei Zhang, Xiaoguang Rui, Rongfang Bie,Early Air Pollution Forecasting as a Service: an Ensemble Learning Approach,IEEE 2017 © 31.00$ 17/7 -0752-5386 -1 -978 DOI 10.1109/ICWS.2017.76

[8] Dixian Zhu, Changjie Cai, Tianbao Yang and Xun Zhou ,A Machine Learning Approach for Air Quality Prediction: Model Regularization and
Optimization:Comput.2018,2,5; doi:10.3390/bdcc2010005

[9] Masoume Asghari Esfandani and Hossein Nematzadeh,Prediction of Air Pollution in Tehran based on Evolutionary Models,Vol 8(35), DOI: 10.17485/ijst/2015/v8i35/56235, December 2015

[10] Surajit Chattopadhyay and Goutami Bandyopadhyay , Artificial Neural Network to predict mean monthly total ozone in Arosa, Switzerland
,10.1080/01431160701250440

[11] Ebrahim Sahafizadeh and Esmail Ahmadi, Prediction of Air Pollution of Boushehr City Using Data Mining, 978-0-7695-3937-9/09 $26.00 © 2009 IEEE DOI
10.1109/ICECS.2009.18