International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT217331

Air Pollution Evaluation by Combining Stationary, Smart Mobile Pollution Monitoring and Data-Driven Modelling

A. Shifa*[1], Dr. S. Rathi[2]

*[1]ME Student, Department of Computer Science and Engineering, Government College of Technology, Coimbatore, Tamil Nadu, India

[2]Professor, Department of Computer Science and Engineering, Government College of Technology, Coimbatore, Tamil Nadu, India

Article Info
Volume 7, Issue 3
Page Number : 222-227

Publication Issue :
May-June-2021

Article History
Accepted : 15 May 2021
Published : 22 May 2021

ABSTRACT

Air pollution has become a major issue in large cities because increasing traffic, industrialization and it becomes more difficult to manage due to its hazardous effects on the human health and many air pollution-triggering factors. This paper puts forth a machine learning approach to evaluate the accuracy and potential of such mobile generated information for prediction of air pollution. Temperature, wind, humidity play a vital role in influencing the pollution dispersion and accumulation, majorly influencing the prediction of pollution levels. Thus, this paper includes the atmospheric condition information registered throughout the study period in order to understand the influence of these factors on air pollution monitoring. Data driven modelling is an efficient way of extracting valuable information from generated data sets, however it is less efficient when the data is incomplete or contains inaccuracies. This modelling approach has true potential for real time operations because it can detect non-linear spatial relationships between sensing units and could aggregate results for regional investigation. Neural networks comparatively showed good capability in air quality prediction than support vector regression.

Keywords : Air Pollution, AQI value, Neural Networks, Support Vector Regression, R2 value

I. INTRODUCTION

Addressing air pollution problems in growing urban cities has become a serious downside due to ever-increasing traffic in densely inhabited urban areas, extended industrialization, high-energy consumption, skimpy resources for monitoring and various issues in shaping custom-made policies. The challenge of managing air pollution becomes tougher because of its dangerous effects on public health and the multitude of air pollution triggering factors. Therefore, numerous studies in recent years are concentrating on evaluating the impact of bad air quality on citizens. This is done by moving away from traditional monitoring stations which are normally placed in high altitude locations across cities, towards outdoor and easy deployable air quality monitoring units, such as mobile sensors installed on cars, bikes or even carried by hand during daily travelling. This new form of collective approach for monitoring air quality brings numerous advantages in terms of real-time pollution measurement and hot-spot identification, however conjointly comes with various challenges due to the amount of information generated and its accuracy. Therefore, there is a true challenge of not only shifting towards a mobile air pollution-monitoring paradigm (and selecting the best-adapted sensing units) but also in modelling efficiently the data generated by all these mobile sensing units.

II. RELATED WORK

The traditional methods for air quality evaluation use mathematical and statistical techniques. In these techniques, initially a physical model design is created and data is coded with mathematical equations. But such methods suffer from discrepancies like: limited accuracy due to inability in predicting the extreme points i.e. the pollution maximum and minimum, cut-offs cannot be achieved, they use inefficient approach for more acceptable output prediction, the presence of complex mathematical calculations and equal treatment to the old data and new data.

However, with the advancement in technology and research, alternatives to traditional ways are projected which use big-data and machine learning approaches. In recent times, several researchers have developed or used big data analytics models and machine learning based models to conduct air quality analysis to realise better accuracy in evaluation and prediction.

Machine learning algorithms are best suited for air quality prediction since it is the branch of computer science, which makes computers capable of performing a task without any explicit programming. Earlier studies focus on classification of air quality evaluation using various machine-learning algorithms. Most of these use different scientific methods, approaches and ML models to predict air quality.

The main objective of this paper is to fit a regression model on the training set and evaluate the model performance using the Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2). Two regression models such as support vector regression and multi-layer perception (Artificial Neural Networks) are evaluated based on the performance metrics mentioned to find the optimum algorithm, which efficiently deals with non-linear spatial relationships among information. The goal is to build collective data-driven predictions for insuring continuous real-time situation awareness.

III. IMPLEMENTATION

A. AIR QUALITY EVALUATION PARAMETER

There is one important parameter known as air quality index (AQI) that quantifies air quality as shown in Table 1. It is a number used by government agencies to communicate to the public how impure the air is presently or how polluted it is forecasted to become. As the AQI value increases, proportionally large percentage of the population is likely to be exposed, and people might experience increasingly severe health issues. Different countries have their specific air quality indices, corresponding to different national air quality standards.

TABLE I. AQI CLASSIFICATION

AQI	Air Pollution Level
0-50	Excellent
51-100	Good
101-150	Lightly Polluted
151-200	Moderately Polluted
201-300	Heavily Polluted
300+	Severely Polluted

B. DATA PREPARATION

The first stage of module implementation is dataset collection. The dataset consists of about approximately 44,000 entries collected over a region for a particular period of time in .csv format. The entries constitutes of the pollution measured, denoted by AQI (Air Quality Index) and a wide range of environmental factors such as dew, temperature, pressure, wind speed, wind direction, snow and rain. Then the data set is divided into training and testing sets. In this implementation, training data is about 35,000 approximately and the remaining is used for testing.

C. MACHINE LEARNING PREDICTION MODELS

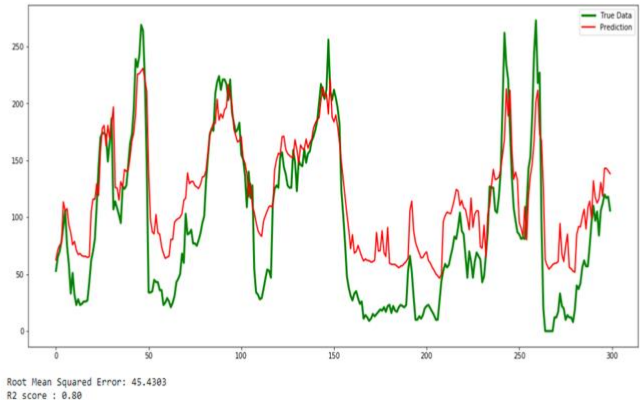
The data collected by mobile sensing unit can be used learn patterns of air pollution evolution, particularly when being used in specific urban locations. When passing through a polluted area, if the pattern analysis detects anomalies and historical high pollution levels, the mobile unit can release alarms to the user to avoid the particular area. In order for this to happen, the information collected by the mobile unit needs to be accurate enough and has to contain enough information that could be used for predicting air pollution depending on location environmental conditions.

1. SUPPORT VECTOR REGRESSION

SVMs are a collection of inter-linked supervised learning methods used for classification and regression, and they are known for being universal approximators of any multivariate function to any desired degree of accuracy. The SVM was originally formed for classification, and was later generalized to solve regression problems. This method is known as support vector regression (SVR). This SVR fit captures the main idea of statistical learning theory to get a good forecasting of the dependence between the main determinants of pollution.

In SVR implementation, the training is done for data collected for about 35000 hours approximately and the remaining is considered to be testing data. The variables in the dataset file are split into separate fields and the support vector regression algorithm is applied to train and to predict the remaining values.

TRUE VS PREDICTED VALUE FOR SVR



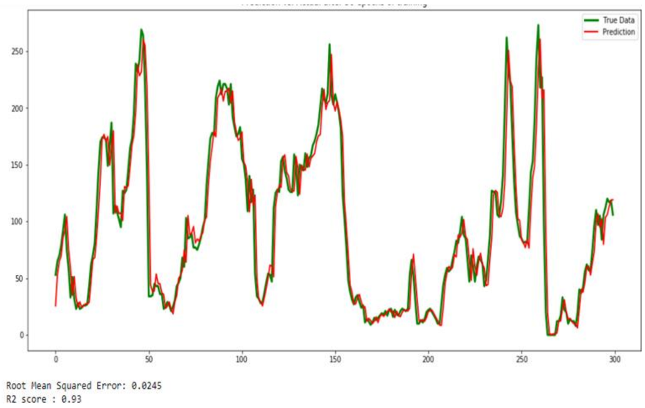
2. ARTIFICIAL NEURAL NETWORK (MLP REGRESSION)

MLP stands for a multilayer perception, which is a well-known class of Artificial Neural Network (ANN). Moreover, MLP consists of multiple layers of perceptrons or at least three layers of nodes specifically input layer, hidden layer, and output layer. Artificial neural Network model tries to simulate the structures and networks inside the human brain. The architecture of neural networks comprises of nodes that generate a signal or remain silent as per a sigmoid activation function in many cases. ANNs are trained with a training set of inputs and determined output data. For training, the edge weights are manipulated to minimise the training error.

In ANN implementation, a feed forward multi-perceptron network is used consisting of 10 input nodes, one hidden layers of 5 nodes respectively, and one output node. Similar to SVR implementation, the training is done for approximately 35000 datas and the testing is done for the remaining entries.

The implementation of both the regression algorithms shows that Neural networks are comparatively showing good capability in prediction air pollution than support vector regression.

TRUE VS PREDICTED VALUE FOR ANN



D. PERFORMANCE CRITERIA

Some of the statistical evaluations are used to evaluate the model performance such as Root Mean Square Error (RMSE) and coefficient of determination (R^2). The criteria formulas are shown below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \hat{x}_i)^2}{m}}$$

where, m is the number of observations, x_i is the actual value and x_i^ is the predicted value.

$$R^2 = \left[\frac{1}{M}\frac{\sum_{j=1}^{M}\left[\left(Y_j - \bar{Y}\right)\left(X_j - \bar{X}\right)\right]}{\sigma_y \sigma_x}\right]^2$$

Where, M is the number of observations, σ_x is the standard deviation of the observation X, σ_y is the standard deviation of Y, *Xj* is the observed values, $\bar{X}$ is the mean of the observed values, *Yj* is the calculated values, and $\bar{Y}$ is the mean of the calculated values.

E. PERFORMANCE INTERPRETATION

COMPARISON BETWEEN PERFORMANCE OF SVR AND ANN

ML Algorithms	Evaluation Metric	Score
Support Vector Regression	R^2 Score	0.80
Multi-Layer Regression (Artificial Neural Networks)	R^2 Score	0.93

IV. CONCLUSION AND FUTURE ENHANCEMENTS

In this paper, support vector regression and artificial neural network machine learning algorithms are implemented to predict the air pollution with the various environmental factors under consideration. The coefficient of determination evaluation for these two algorithms showed that the prediction accuracy for neural networks is increased by about 13% than that of support vector regression. The increase in performance is due to the capability of neural networks to deal with non-linear spatial relationships in data.

There is a lack of solutions proposing both real-life air quality monitoring at human level and data-driven prediction approaches for situation awareness and real-time alert generation. The accuracy that can be achieved through the proposed algorithm can be extended to feed an application like Google Maps. Instead of detecting the traffic and suggesting a different route, this can warn the pedestrians and cycle-riders to take a different route due to more pollution in a particular area.

V. REFERENCES

[1]. Adriana Simona Mihaita, Laurent Dupont, Olivier Chery, Mauricio Camargo ,Chen Cai: Evaluating air quality by combining stationary,

smart mobile pollution monitoring and data-driven modeling, https://doi.org/10.1016/j.jclepro.2019.02.179

[2]. Chris C. Lim, Ho Kim, M.J. Ruzmyn Vilcassim, George D. Thurston, Terry Gordon, Lung-Chi Chen, Kiyoung Lee, Michael Heimbinder, Sun-Young Kim: Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea, https://doi.org/10.1016/j.envint.2019.105022

[3]. Pooja Bhalgat, Sejal Pitale, Sachin Bhoite: Air Quality Prediction using Machine Learning Algorithms. In: International Journal of Computer Applications Technology and Research Volume 8–Issue 09, 367-370, 2019, ISSN:-2319–8656

[4]. Doreswamy, Harishkumar K S1, Yogesh KM, Ibrahim Gad : Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models. In: Third International Conference on Computing and Network Communications (CoCoNet'19) Procedia Computer Science 171 (2020) 2057–2066

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

[6]. A.Suárez Sánchez, P.J.García Nieto, P.Riesgo Fernández, J.J.del Coz Díaz, F.J.Iglesias-Rodríguez: Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain) https://doi.org/10.1016/j.mcm.2011.04.017

[7]. Bing-Chun Liu, Arihant Binaykia, Pei-Chann Chang, Manoj Kumar Tiwari, Cheng-Chin Tsao: Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR) https://doi.org/10.1371/journal.pone.0179763

[8]. Heidar Maleki, Armin Sorooshian, Gholamreza Goudarzi, Zeynab Baboli, Yaser Tahmasebi Birgani,Mojtaba Rahmati : Air pollution prediction by using an artificial neural network model. In: Clean Technologies and Environmental Policy volume 21, pages1341–1352(2019)

[9]. V. M. Niharika and P. S. Rao, "A survey on air quality forecasting techniques, "International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp.103-107, 2014.

[10]. E. Kalapanidas and N. Avouris, "Applying machine learning techniques in air quality prediction," in Proc. ACAI, vol. 99, September 1999.

[11]. D. J. Nowak, D. E. Crane, and J. C. Stevens, "Air pollution removal by urban trees and shrubs in the United States," Urban Forestry & Urban Greening, vol. 4, no. 3, pp. 115-123, 2006.

[12]. T. Chiwewe and J. Ditsela, "Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations," presented at 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), IEEE, 2016.

[13]. Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data," in Proc. the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2267-2276, August 10, 2015.

[14]. J. A. Engel-Coxa, C. H. Hollomanb, B. W. Coutantb, and R. M. Hoffc, "Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality," Atmospheric Environment, vol. 38, issue 16, pp. 2495–2509, May 2004.

[15]. J. Y. Zhu, C. Sun, and V. Li, "Granger-Causality-based air quality estimation with spatio-temporal (ST) heterogeneous big data," presented at 2015 IEEE Conference on

Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2015.

[16]. C. J. Wong, M. Z. MatJafri, K. Abdullah, H.S. Lim, and K. L. Low, “Temporal air quality monitoring using surveillance camera,” presented at IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2007.

[17]. S. Y. Muhammad, M. Makhtar, A. Rozaimee, A. Abdul, and A. A. Jamal, “Classification model for air quality using machine learning techniques,” International Journal of Software Engineering and Its Applications, pp. 45-52, 2015.

[18]. A. Sarkar and P. Pandey, “River water quality modelling using artificial neural network technique,” Aquatic Procedia, vol. 4, pp. 1070-1077, 2015.

[19]. E. Kalapanidas and N. Avouris, “Applying machine learning techniques in air quality prediction,” Sept. 1999.

[20]. H. Zhao, J. Zhang, K. Wang, et al., “A GA-ANN model for air quality predicting,” IEEE, Taiwan, 10 Jan. 2011.

[21]. R. Yu, Y. Yang, L. Yang, G. Han, and, O. A. Move, “RAQ–A random forest approach for predicting air quality in urban sensing systems,” Sensors, vol. 16, no. 1, p. 86, 2016.

[22]. S. Deleawe, J. Kusznir, B. Lamb, and D. J. Cook, “Predicting air quality in smart environments,” J Ambient Intell Smart Environ., pp. 145-152, 2010.

[23]. W. F. Ip, C. M. Vong, J. Y. Yang, and P. K. Wong, “Least squares support vector prediction for daily atmospheric pollutant level,” in Proc. 2010 IEEE/ACIS 9th International Conference on Computer and Information Science (ICIS), pp. 23-28, IEEE., August 2010.

Cite this Article

A. Shifa, Dr. S. Rathi, "Air Pollution Evaluation by Combining Stationary, Smart Mobile Pollution Monitoring and Data-Driven Modelling", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 222-227, May-June 2021. Available at doi : https://doi.org/10.32628/CSEIT217331

Journal URL : https://ijsrcseit.com/CSEIT217331