International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 2545-2549 www.ijaem.net ISSN: 2395-5252

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*

Submitted: 10-06-2021 Revised: 20-06-2021 Accepted: 23-06-2021

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 for prediction of air pollution. Levenberg-Marquardt algorithm is extremely dependent on the initial guess for the network parameters. Depending on the initial weights of the network, the algorithm may converge to local minima or not converge at all. Bayesian regularization expands the cost function to search not only for the minimal error, but also for the minimal error using the minimal weights. Thus, this project proposes an incorporation of LM algorithm with BR algorithm to overcome the disadvantages faced when the individual algorithm is used. 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. The incorporation of Bayesian Regularization with Levenberg-Marquardt neural network algorithm gives better results than the individual algorithms.

KEYWORDS: Air Pollution, NO_2, Neural Networks, Levenberg-Marquardt algorithm, Bayesian Regularisation algorithm, R^2 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 Levenberg-Marquardt Algorithm and Bayesian Regularization Algorithm (Artificial Neural Network algorithms) 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. NO_2 AS A AIR POLLUTANT

Nitrogen dioxide (NO_2) is one of the nitrogen oxides (NO_x), a group of air pollutants produced from combustion processes. In urban outdoor air, the presence of NO_2 is mainly due to traffic. Nitric oxide (NO), which is emitted by motor vehicles or other combustion processes, combines with oxygen in the atmosphere, producing NO_2. Mainly unvented heaters and gas stoves produce indoor NO2. NO_2 and other nitrogen oxides are also precursors for a number of harmful secondary air pollutants such as ozone and particulate matter, and play a role in the formation of acid rain. Exposure to NO_2 may affect health independently of any effects of other pollutants.

TABLE I
INDEX SCALE FOR NO_2

Value(µg/m^3)	Index	Air Pollution Level
0-29	1	Very Good
30-54	2	Very Good
55-84	3	Good
85-109	4	Good
110-134	5	Medium
135-164	6	Medium
165-199	7	Poor
200-274	8	Poor
275-399	9	Bad
>=400	10	Very Bad

B. DATA PREPARATION

The first stage of module implementation is dataset collection. The dataset consists of about approximately 9,334 entries collected over a region for a particular period of time in .csv format. The entries constitutes of the concentration of the air pollutant (NO_2) measured 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 6,500 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. LEVENBERG-MARQUARDT ALGORITHM

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

Levenberg's main contribution to the method was the introduction of the damping factor **λ**. This value is summed to every member of the approximate Hessian diagonal before the system is solved for the gradient. Typically, **λ** would start as a small value such as 0.1.

The training problem can be considered as a general function optimization problem, with the adjustable parameters being the weights and biases of the network.

$(J^tJ + \lambda I)\delta = J^tE$

Jacobian J is a matrix of all first-order partial derivatives of a vector-valued function. In the neural network case, it is a **N**-by-**W** matrix, where **N** is the number of entries in our training set and **W** is the total number of parameters (weights + biases) of our network.

After the equation is solved, the weights **w** are updated using **δ** and network errors for each entry in the training set are recalculated. If the new sum of squared errors has decreased, **λ** is decreased and the iteration ends. If it has not, then the new weights are discarded and the method is repeated with a higher value for **λ**.

2. BAYESIAN REGULARIZATION ALGORITHM

Bayesian regularization expands the cost function to search not only for the minimal error, but for the minimal error using the minimal weights. It works by introducing two Bayesian hyper parameters, alpha and beta, to tell which direction (minimal error or minimal weights) the learning process must seek.

The cost function will then become:

$C(k) = \beta * E_d + \alpha * E_w$

where:

E_d is the sum of squared errors, and

E_w is the sum of squared weights

Update the Bayesian hyper parameters using MacKay's or Poland's formulae:

gamma = W – (alpha * $\mathrm{tr}(H^{-1})$)

beta = (N – gamma) / 2.0 * E_d

alpha = W / (2.0 * E_w + tr(H-1)) [modified Poland's update], or

alpha = gamma / (2.0 * E_w) [original MacKay's update],

where: W is the number of network parameters (number of weights and biases), N is the number of entries in the training set and $\mathrm{tr}(H^{-1})$ is the trace of the inverse Hessian matrix.

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 $x_i^{\wedge}$ is the predicted value.

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

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 PERFORMANCES OF ANN ALGORITHMS

ML Algorithms	Evaluation Metric	Score
Multi-Layer Regression (Artificial Neural Networks	R^2 Score	0.67
Levenberg-Marquardt Algorithm	R^2 Score	0.75
Levenberg-Marquardt Algorithm with Bayesian Regularisation	R^2 Score	0.82

IV. CONCLUSION AND FUTURE ENHANCEMENTS

In this paper, 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 when the Bayesian Regularization is used along with Levenberg-Marquardt algorithm. The increase in performance is due to the capability of regularization to reduce (or eliminate) the need for testing different number of hidden neurons for a problem. A third variable indicates the number of effective weights being used by the network, thus giving an indication on how complex the network should be.

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.

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.

[24] Annunziata Faustini, Regula Rapp, Francesco Forastiere, "Nitrogen Dioxide and Mortality: Review and Meta-Analysis of Long Term Studies" European Respiratory Journal 2014 44: 744-753;
DOI: 10.1183/09031936.00114713

[25] Annunziata Faustini, Regula Rapp, Francesco Forastiere, "Nitrogen Dioxide and Mortality: Review and Meta-Analysis of Long Term Studies" European Respiratory Journal 2014 44: 744-753;
DOI: 10.1183/09031936.00114713

[26] A. Payal, C. S. Rai and B. V. R. Reddy, "Comparative analysis of Bayesian regularization and Levenberg-Marquardt training algorithm for localization in wireless sensor network," 2013 15th International Conference on Advanced Communications Technology (ICACT), 2013, pp. 191-194.

[27] Mehmet Pakdemirli, "Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data", Math. Comput. Appl. 2016, 21, 20; doi:10.3390/mca21020020

[28] Yasemin Gültepe, Ayşe Mine Duru, "Daily SO2 Air Pollution Prediction with the use of Artificial Neural Network Models", International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 34, December 2018

[29] Leu, Fang-Yie & Ho, Jia-Sheng, "Air Pollution Source Identification by Using Neural Network with Bayesian Optimization (2020)", DOI: 10.1007/978-3-030-22263-5_49

[30] Zhu D, Cai C, Yang T, Zhou X, "A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization", Big Data Cogn. Comput. 2018, 2, 5. DOI: 10.3390/bdcc2010005