© 2018 IJSRST | Volume 4 | Issue 3 | Print ISSN : 2395-6011 | Online ISSN: 2395-602X

National Conference on Advances in Engineering and Applied Science (NCAEAS)
29th January 2018
Organized by : Anjuman College of Engineering and Technology (ACET) Nagpur, Maharashtra, India, In association with
International Journal of Scientific Research in Science and Technology

A Review on Urban Air Pollution Monitoring System with Forecasting Model

Roshni Jeswani[1], Mohd. Tahseenul Hasan[2]
[1]Student, Electronics and Communication, RTMNU, ACET, Nagpur, Maharashtra, India
[2]Assistant Professor, Electronics and Telecommunication, RTMNU, ACET, Nagpur, Maharashtra, India

ABSTRACT

In this paper the system for monitoring and forecasting urban air pollution is presented. The system uses low-cost air-quality monitoring motes that are equipped with an array of gaseous and meteorological sensors. These mote wirelessly communicate to an intelligent sensing platform that consists of several modules. These modules are responsible for receiving and storing the data, preprocessing and converting the data into useful information, forecasting the pollutants based on historical information, and finally presenting the acquired information through different channels, such a mobile application, Web portal, and short message service. This paper focus is on the monitoring system and its forecasting module. Three machine learning (ML) algorithms are investigated to build accurate forecasting models for one-step and multi-step ahead of concentrations of ground-level ozone (O3), nitrogen di oxide (NO2), and sulfur dioxide (SO2), carbon dioxide (CO2). Rapid urbanization and industrialization has result in a sustained degradation of environmental quality parameters. It is important to keep track of various environmental pollution indices so that realistic models can be developed and relevant public policies can be created. Traditional methods for air pollution measurement are expensive and have a spatial constraint. With these limitations, air pollution monitoring broader area is not feasible. We used the modern low-cost sensors in conjunction with wireless sensor network (WAN) creates an opportunity to collect real time data from different locations and provide detailed pollution map. The main aim of this project is to develop a low cost multi-sensor node for air pollution measurement. The outcome of this paper can be significantly useful for alarming applications in areas with high air pollution levels.

Keywords: Air pollution monitoring system, forecasting, machine learning approach.

I. INTRODUCTION

Air quality is one of the major environmental problem for people, because of its direct influence on the state of human health. In this application both the interface and the computation complexity required the use of new and powerful computer paradigms. In the use of expert systems and agents has supplied a valid solution to improve complex acquiring-tasks about suitable parameters. The interest and the attention devoted to the environmental thematic, to the monitoring and to control activities about air quality, are growing quickly Environmental National Agencies are imposing the implementation of environmental monitoring stations located in a wide geographic area, in order to supply representative real data related to the atmospheric pollution processes characterize the causes determining the pollution phenomena [18].Across the world, increasing population and rapid industrialization has caused significant environmental degradation. It ranges across air, water, noise and land pollution. Carbon Monoxide (CO), Carbon Dioxide (CO2),Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), Particulate Matter (PM), Lead (pb) ,Ammonia (NH3), Ground level Ozone (O3) are the primary cause of air pollution. Development of air pollution monitoring system will be beneficial to control and measure pollution related parameters. Conventional strategies for measurement of air pollution parameters are more precise yet expensive and restricted to spatial area, it is not possible to deploy measurement instruments in large number. In remote locations i.e. glaciers where less or no network connectivity is found then data communication and collection becomes major issue. Because of the communication issue, data needs to be collected manually at fixed location which is time consuming [14]. Many studies on human health have concluded that environmental stress is a major factor for morbidity and has a negative impact on the quality of life especially in urban areas (e.g.).One of the major challenges in these studies is to obtain or estimate high resolution (spatial and temporal) air quality data to be able to analyze the correlation between health and the exact air to which people are exposed. Among all the airborne pollutants (Sox, NOx ,CO, NH3 O3, etc.). Recently there has been a growing attention to study particulate matters due to their significant adverse impact on human health. In urban environments, this measure is closely linked to urban traffic conditions [8]. Recent development of electronics has realized the vision of using wireless communication in devices used for monitoring wide range of real life parameters, such as temperature, pressure, and air pollution. These devices send their measurements wirelessly to a database hosted on a remote server for further processing and analysis. The concept of using small size, inexpensive motes that wirelessly communicate their air pollution measurements has been widely studied and implemented [2].

II. AIR POLLUTION MONITORING SYSTEM

Air pollution attracts extensive attention worldwide due to its tremendous impacts on human health, global environment and economy. Conventional monitoring systems have been deployed to provide authorized information for urban management and environmental improvement. These systems have extremely low spatial and temporal resolutions and are inadequate for monitoring personal and acute exposures to air pollutants [1].

This paper describes the implementation and evaluation of an Air Pollution Monitoring System (APMS) for monitoring the air pollution using a WSN; this also constitutes the main contribution of this work. More specifically, in March 2015, a dense sensor network composed of 10 sensor nodes has been installed in a vibrant part of the city, covering an area of approximately 1 km2, equipped with sensors for measuring temperature, humidity, noise, light, CO2, CO, NO2, O3 and PM10. The data from the sensors is collected wirelessly every 30 min through a gateway and stored in a database. A web interface has also been designed for providing easy access and viewing of the sensor data. In the paper, we describe the various parts of the system in terms software using MATLAB. We detail the installation procedure with particular emphasis on the method used for calibrating the sensors. Finally, we offer some preliminary findings in comparing the sensor readings between the different nodes. Our primary purpose in this paper is to share our experiences in applying WSNs technology for the problem of air pollution monitoring in an urban environment. Towards this end, we explain the various design choices involved as well as the problems and difficulties we encountered along the way. “Air Cloud System” is proposed, composed of personal low-cost internet-connected particulate matter (PM) sensor monitors and an air quality modelling engine providing accurate device calibration and fine-granularity estimation based on GPS-location. In Mauritus, investigates the use of a Wireless Sensor Network for air pollution monitoring. It also describes data aggregation algorithms for eliminating duplicates and summarising data into a simpler form. In, the implementation of a low cost air quality monitoring solution in Lahore, Pakistan is described, as part of the VIEW (Volunteer Internet-based Environment Watch) project. The feasibility of low cost electromechanical sensors for monitoring urban air quality is further demonstrated. The authors provide evidence for the performance of electrochemical sensors in the parts-per-billion (ppb) level for gas species (NO, NO2 and CO) and outline results from deployments of static networks of such sensor nodes and mobile networks for quantifying personal exposure in Cambridge, UK. In Japan, describes the development and calibration of a gas sensor system (NO2) to be used as a sensing node to form a dense real-time environmental monitoring network.

III. FORECASTING

Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends. A common place example might be estimation of some variable of interest at some specified future date. Prediction is a similar, but more general term. Both might refer to formal statistical methods employing time series , cross-sectional or longitudinal data, or alternatively to less formal judgmental methods. Usage can differ between areas of application: for example, in hydrology the terms "forecast" and "forecasting" are sometimes reserved for estimates of values at certain specific future times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period. Risk and uncertainty are central to forecasting and prediction; it is generally considered good practice to indicate the degree of uncertainty attaching to forecasts. In any case, the data must be up to date in order for the forecast to be as accurate as possible. In some cases the data used to predict the independent variable is itself forecasted.

IV. MACHINE LEARNING APPROACH

ML involves computational methods that improve the performance of mechanizing the acquisition of knowledge from experience. Machines learn from complex data to be able to solve problems, answer questions and be more intelligent. One of the tasks that highly involve learning is forecasting, in which the forecasting model is built through training from data that is generally nonlinear in the case of air quality. Therefore, approached based on linear modeling may not be suitable for such data . After training, the model is ready to predict unseen data and hence can answer the forecasting question, for example: “What will be the next hour value of NO2 gas concentration in air?” Specifically, we aim to accurately predict concentrations of O3, NO2, and SO2 as they are considered to be the most harmful gases. Before employing the nonlinear modelling methods, the nonlinear structure of the data is verified for all gases. Brocke-Decherte-Scheinkman

(BDS) is used. The BDS statistic, ωm,n(ε), is computed and the nonlinearity in data is verified if the null hypothesis of linearity is rejected at the 5% significance level. This condition is applicable if |ωm,n(ε)| > 1.96. If the time series data comprises more than 7500 observations, as in our case, the BDS statistic is derived in terms of the correlation integral, cm,n(ε), using the formula:

$$\omega_{m,n}(\varepsilon) = \sqrt{n}\frac{c_{m,n}(\varepsilon) - c_{1,n}^{m}(\varepsilon)}{\sigma_{m,n}(\varepsilon)}$$

Where n is the sample size (8832 observations for all gases, in our case), m is the embedding dimension and it takes a discrete value in the range [2]–[5] at the big sample size σ is the standard deviation of time series and ε takes a recommended value in the range from 0.5σ to 2σ based on the assumption that samples have normal or near-normal distribution. For our samples, ωm ,n(ε) is computed, by equation (1), for

SO_2, NO_2, and O_3 are found to take values in the ranges of [207.36-266.12], [140.72-176.42], and [191.43 -224.35]. These values are extremely greater than 1.96, which reveal the sharp nonlinearity in data.

Our methodology consists of the following steps:

1. Data Preprocessing: Data received from the sensors will be subjected to preprocessing to remove outliers and anomalies and the data will be prepared in the format acceptable by the ML learning algorithm.

2. Feature Engineering: This step is concerned with selecting the features to be included in the prediction process along with each target gas, such as temperature, humidity, and day of the week.

3. Time Windowing: This is a fundamental task with time series forecasting, in which a number of time-lagged features for each input attribute is generated in order clarify the time dependency between consecutive data points. Window size, step size and horizon are key parameters that control time windowing. Window size is the number of generated features (i.e., generating multi-dimensional vectors) from the single-dimensional data. Step size is the number of instances between windows Horizon is the number of steps in the future to before casted.

4. Building Forecasting Models: Models will be designed to predict future values based on historical data using ML algorithm.

The process encompassing the above steps to construct and apply ML-based models for predicting values of unseen target data is depicted in Figure 1.

In training, data with known target values are collected; a subset of feature is selected, and then used to construct a forecasting model. There are many subsets of features selected and various ML algorithms used; therefore, there are various predictors that can be trained.

In testing, the produced models from the training phase are validated and evaluated. Several methods are used in model validation, such as different sliding windows, in which two windows are used for training and testing and each has its own size, step size, and horizon. This validation method guarantees that instances used for testing are not known before to the model through training, hence reliable performance measures are calculated such as prediction trend accuracy (PTA) and root mean square error (RMSE). PTA is a time series measurement of how close is the predicted data trend from the trend of the actual data

First, actual trend (AT) and predicted trend (PT) are calculated as:

AT = Label [i] – Label [i – horizon]

PT = Predicted [i] – Label [i – horizon]

Where Label is the target feature, i is the instance number, and horizon is the number of steps forecasted in the future.

Trends are then multiplied by each other. If the result is greater than or equal to zero, then the actual and predicted trends have the same sign, hence have the same trend, so a counter is incremented. This process is repeated for all data, and finally, the counter is divided by the total number of instances. RMSE is a common performance metric in model evaluation and it is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y - \hat{y})^2)}$$

where n is the number of instances, y is the actual value of the target feature and ˆy is the predicted value of it.

Normalized RMSE (NRMSE) is used to compare the performance of different models predicting different target variables and it is calculated as:

$$NRMSE = \frac{RMSE}{(y_{max} - y_{min})}$$

Where ymax and ymin are the maximum and minimum values of collected data.

In the deployment phase, the best model and features will be used to process unseen data and produce prediction results. The model performance is kept on check to validate its prediction results.

Practically, and especially in changing environments, the process of training, testing, and deployment are periodically repeated to maintain high accuracy of results. Moreover, this iterative process can be performed to improve performance of the models as historical data become increasingly available. There exists a plethora of algorithms to build ML-based forecasting models that may behave differently to the given data [2].

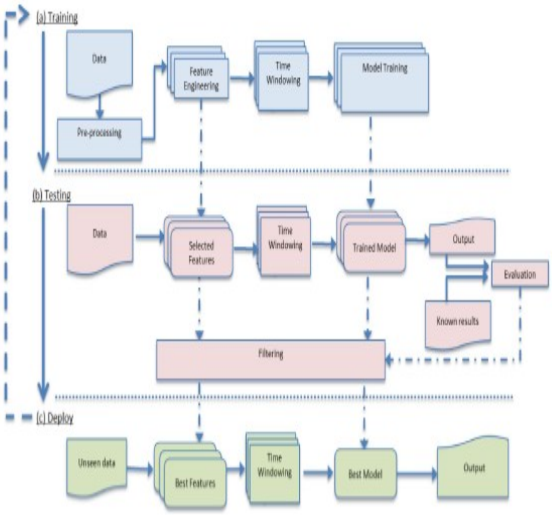

Fig. 1. Iterative process of constructing and applying ML-based prediction models.

V. APPLICATION AND FUTURE SCOPE

It is significantly useful for alarming applications in areas with high air pollution levels and the used regression model to predict air pollution concentration for health exposure studies. In future it is advantageous because it will identify pollution sources and predicting urban air quality using MLA methods. Its low cost sensors in conjunction wireless sensor network creates an opportunity to collect real time data from different location and provide detailed pollution map.

VI. CONCLUSION

Air pollution is the major problem of today's society that affects both the environment and human health. In this paper we focus on monitoring system and forecasting model for the daily forecast of pollutant concentration. This paper has proposed a new time series based on forecasting model to predict air quality by maximum O3 concentration. From the air quality pollutes main is caused by O3.future search can test whether the air quality have influenced the nearby human health and the security under high-tech operation. These models predict 1, 8, 12, and 24 hours ahead of concentration values.

VII. REFERENCES

[1]. Khaled Bashir Shaban, Senior Member, IEEE, Abdullah Kadri, Member, "Urban air pollution monitoring system with forecasting model" IEEE, and Eman Rez VOL.16,NO.8,APRIL 15,2016,IEEE .

[2]. Wei-Ying Yi, Kwong-Sak Leung, Yee Leung, Mei-Ling Meng, and Terrence Mak, "Modular Sensor System (MSS) for Urban Air Pollution Monitoring", UK, 978-1-4799-8287-5/16 2016 IEEE.

[3]. Theofylaktos Pieri and Michalis P. Michaelides "Air Pollution Monitoring in Lemesos using a Wireless Sensor Network", 978-1-5090-0058-6/16 2016 ,IEEE

[4]. Mihaela Oprea "A knowledge based approach for PM2.5 air pollution effects analysis", 978-1-4673-9910-4/16 2016 IEEE

[5]. Mihaela Oprea1, Cornel Ianache1, Sanda Florentina Mihalache1, Elia Georgiana Dragomir, VALAHIA "On the Development of an Intelligent System for Particulate Matter Air Pollution Monitoring, Analysis and Forecasting in Urban Regions", 978-1-4799-8481-7/15 2015 IEEE.

[6]. Abdul Hadi Nograles H.1, Agbay, Christopher Paolo D.2, Flores, Ian Steven L. 2, Linsangan, Manuel Jr. A. 2, Salonga, John Bethany C. 2 "Low Cost Internet Based Wireless Sensor Network for Air Pollution Monitoring using Zigbee Module, ISBN: 978-1-4799-3724-0/14 2014 IEEE.

[7]. Le Dinh Minh1, Sang Boem Lim1, Jung-Hun Woo1, HyungSeok Kim2 "A Data Transferring Model for Micro-scale Urban Air Quality Management System", 978-0-7695-3687-3/09 2009 IEEE.

[8]. Ali Marjovi, Adrian Arfire and Alcherio Martinoli "High Resolution Air Pollution Maps in Urban Environments Using Mobile Sensor Networks" 978-1-4799-8856-3/15 2015 IEEE.

[9]. Hiroaki Kuze,1 Yutaro Goto,1 Yusaku Mabuchi,1 Hayato Saitoh,1 Ilham Alimuddin,1 Gerry bagtasa,1 Ippei Harada,2 Toshihiko Ishibashi,3 Takuma Tsujimoto,3 Shumpei Kameyama 4 "URBAN AIR POLLUTION MONITORING USING DIFFERNTIAL OPTICAL ABSORPTION SPECTROSCOPY (DOAS) AND WIND LIDAR", 978-1-4673-1159-5/12 2012 IEEE.

[10]. Suraj Hebbar', Karuppasamy V',Kiran G K', Amit Kumar', A. S Ashwini Kumari', Raghavendra Yasasvi', Amit K. Gupta, Vijay Mishra', Bharadwaj Amrutur2 and Navakanta Bhat, "System Engineering and Deployment of Envirobat An Urban Air Pollution Monitoring Device" 2014 IEEE

[11]. Motohisa FUNABASHI and Koichi HARUNA "A Mathematical Model for Air Pollution Control in Urban Areas" 2007 IEEE.

[12]. CHING-HSUE CHENG "Forecasting of ozone concentration using frequency MA-OWA Model", 978-1-4244-2794-9/09 2009 IEEE.

[13]. Ahmed Boubrima, Walid Bechkit and Hervé Rivano " Optimal WSN Deployment Models for Air Pollution Monitoring" 1536-1276 (c) 2016 IEEE.

[14]. Bhavika Bathiya Sanjay Srivastava Biswajit Mishra "Air Pollution Monitoring Using Wireless Sensor Network" 978-1-5090-3745-2/16 2016 IEEE.

[15]. Wang Xiwen "The Research of Urban air pollution forecast Base on GIS Technology" 978-1-4244-6542-2/ 2010 IEEE.

[16]. F.M. Raimondi, A. Lo Bue, M.C.Vitale "A CNN adaptive model to estimate PM1O monitoring"2006 IEEE.

[17]. A. Amatol, G. Andria2, T. Delvecchio',V. Di Lecce2, A. Guerriero', A.M.L. LanzoIIa', C. Pasqualei, V. piuri3 "Analysis and metrological characterization of the state of atmospheric pollution by means of mathematical models and agents" 0-7803-7783-4/2003 IEEE