

# Air Pollution Monitoring and Prediction using Multi View Hybrid Model

A.Sivasangari, P.Ajitha, K.Indira

**Abstract**— Pollution monitoring system is used to monitor the air pollution throughout the city, which cause pollution over a specified limit. The sensor nodes are attached to the lamp post. The sensors are organized into clusters and form a mesh network of nodes that provide both single hop and multihop connectivity with the base station. The GPS enabled sensor nodes finds location in order to detect the pollution occurring place. A hybrid model is proposed in this work which combines the spatial and temporal features for prediction. This model use the real time air quality information in a city by measuring the pollution information using sensors and data sets.

**Keywords**— Spatial features, Temporal Features, AQI

## 1. INTRODUCTION

IOT integrated air pollution monitoring system with cloud computing and alert system has not been implemented in any institution or any industry across the world so far. Air pollution due to vehicular and industrial emission has become a threat to the living beings. Due to this both indoor and outdoor air quality monitoring in real time has become mandatory. Air quality information can be used in different applications. Air pollution monitoring system will be used to find the pollution free route for cardiac and respiratory patients. Health awareness people could also use this option for pollution free navigation. The proposed work propose a multi view hybrid model that combines the spatial and temporal predictions to predict the air quality with insufficient and inaccurate data. The proposed work infer the real-time and fine-grained air quality information throughout a city using (historical and real-time) air quality measured by the monitor stations and a variety of datasets.

## 2. RELATED WORK

Citysee is an environment monitoring system for measuring carbon emission measurement in city [1]. It integrates the wireless sensor network and cloud computing applications. Data collection is based on multihop node. Every node is attached with an agent that sense the environment. It provides the real time running status of network and node. Adaptive and distributed routing method is followed in that application.

Ke Hu et al [2] designed pollution monitoring system using pollution sensors, GPS module and 4G module. Mobile app can be used to show the information to the users such as location, type of pollutant and current values of pollutant. Inverse distance weighting can be used to estimate

the concentration at the point. But contours are not generated very smoothly. Mobile app is developed for estimating people individual impact in the pollution. It can be used to understand the relationship between air pollution and its impact.

A low cost air quality monitoring system [3] was developed by using gas and metrological sensors. Data received from the sensors are preprocessed and converted into useful information. Machine learning algorithm is used for building the accurate forecasting model for detecting pollution. The performance of the proposed algorithm is evaluated by calculating accuracy and root mean square. It will alert user through either mobile app or web application when pollution level is high .Better forecasting result is getting by using multivariate modeling.

Goncalo et al [4] proposed an indoor monitoring system for ambient assisted living. Xbee is used for transmitting data from sensor nodes to base station. Energy conservation is maintained by applying sleep mode function for sensor nodes. Users able to see the data status in graph. A Novel hybrid model is proposed for measuring the concentration PM value in air [5]. Trajectory based geographical model is developed for measuring the pollution in neighboring sites. It is capable of measuring the atmospheric and geospatial information. Artificial neural network can be used for predicting the daily average concentrations of PM in air.

Katherine et al [6] proposed a methodology for analyzing pollution levels and gain the knowledge about pollutants exposures. Hu ven et al [7] proposed a study method for analyzing acidic features of air pollutants. Topological map is used for identify the positive correlation and negative correlation with pollutants. A lightweight middleware and web interface is designed for viewing the live pollution information. Gas sensors are calibrated using appropriate calibration techniques [8].

## 3. PROPOSED WORK

The sensor unit consists of sensor, microcontroller and battery for power supply. The sensor component consists of three sensors such as gas sensors, particle sensors and thermal sensors. The particle sensor is used to detect the physical contaminant like dust. The thermal sensor ia used to detect the temperature and humidity. The collected data are transferred to sink node. Sink node will forward the data to the base station. The gas sensors monitor the gases like CO<sub>2</sub>, O<sub>2</sub>, CO and NO<sub>2</sub>. The sensor selection is performed by heuristic based greedy algorithm. The sensor can be removed due to poor connection with base station,

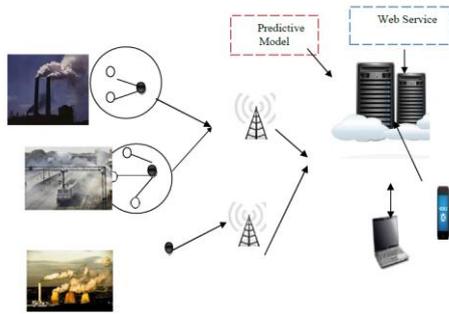
Revised Manuscript Received on April 12, 2019.

**A.Sivasangari**, Asst professor, Sathyabama Institute of Science and Technology, Chennai, T.N, India. (E-mail: sivasangarikavya@gmail.com)

**P.Ajitha**, Asst professor, Sathyabama Institute of Science and Technology, Chennai, T.N, India. (E-mail: hannahgracelyne@gmail.com)

**K.Indira**, Asst professor, Sathyabama Institute of Science and Technology, Chennai, T.N, India. (E-mail: indirakaliannan@gmail.com)

correlation of data with other sensors and a sensor has low energy ie already it is used. The uniform quantization is applied in sensor data to reduce the storage and transmission cost. The following fig 1 shows the proposed architecture for air pollution.



**Fig.1 Proposed Architecture**

Government use an air quality index (AQI) as a number to communicate to the user about polluted air at current time and how it is polluted. Air quality index is measured as the ratio of measured concentration and reference concentration values. A high AQI value indicates the poor air quality and it can affect the health of people. Ambient air pollutant are classified such as oxides of nitrogen, sulphur dioxide, carbon monoxide, particular matters and ground level ozone.

Main factors that affect the concentration of pollution are vehicular and meteorological. The vehicular factors are based on the vehicle count, date and time of the vehicle. The meteorological factors are based on the parameters such as speed of the wind, direction of wind, environment humidity and temperature. Historical information is obtained by using concentration of pollution 1 day before, 1 week before and 1hr before. In a road network, road components are connected with each via graph format.

Metrology will influence the concentration of pollutants. High speed wind disperses the PM10 concentration and high humidity produces a high concentration. High pressure causes the good AQI result. The temporal predictor find the air quality by measuring local meteorology data, past hours values of AQI, features of traffic and weather forecast information of particular place. Traffic flow is considered as one of major issue in air pollution. Traffic features are measures by path of the moving vehicles traversing in the grid in the past hours. The distance between two vehicles are calculated. The spatial predictor measures the spatial neighbor data by using AQI values, wind speed of current station and other stations.

Based on the two separated data set, two separate classifiers are trained. The temporal classifier is based on a support vector machine and gaussian kernel function. It will estimate the temporal transformation quality of air station in the particular location by using temporal features. The spatial classifier is based on artificial neural network(ANN) to model the spatial correlation between air qualities of differentstations.

- 1) The spatial predictor will take the input feed from every station by using aggregated air quality information and metrology information. It will predict

the air quality based on the inputs.

- 2) Extract the AQI  $i$  for past three hours, metrological features  $M_i$  and current time  $t_c$  from time series.
- 3) Derivation between the AQIs of current time  $t_c$  and future time interval  $t_{c+w}$

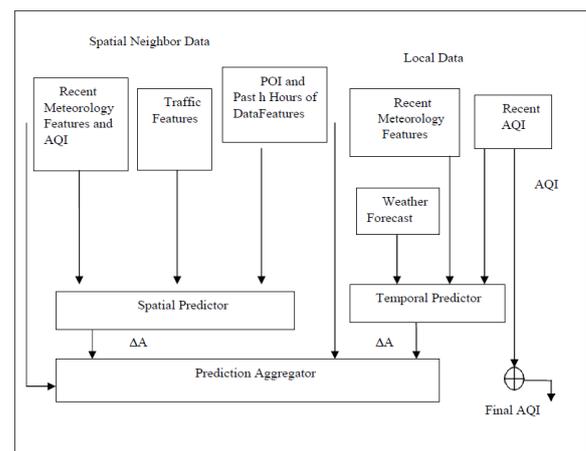
$$\Delta AQI = AQI_{t_c} - AQI_{t_{c+w}} \quad (1)$$

- 4) The regression tree is used to find the speed of that vehicles by employing the prediction generated by spatial and temporal predictors.

Back propagation algorithm is used to train the network. The model performance is kept on check to validate its prediction results. It consists of input layer, hidden layer and output layer. The performance of the network is based on connection weights, transfer function and hidden layer. Sigmoid function calculate can be used to

The proposed algorithm calculates the weights between the hidden layer and output layer. The errors of the network are propagated back to the network. The value of each hidden node is calculated as the activation function of total weighted input of that node. The prediction analysis is described in fig2.

In this phase, the output of the network is compared to the actual correct output and an error is calculated. This error is propagated from the output back to the hidden and input layer and adjusting the weights of each connection in the network layers. This procedure is repeated to minimize the error. The mobile app displays the pollution information such as current location, type of the pollutant, current time and historical data. It requires only minimal input from the user side are needed to start the mobile application.



**Fig. 2 Prediction Analysis**

#### 4. PERFORMANCE ANALYSIS & RESULTS

Pollution data were partitioned into two sets such as 85% for training and 15% for testing. Normalization can be performed in the dataset for producing unique values. The proposed work is to examine the effectiveness of the hybrid model for the prediction of pollutant concentrations. Spatial prediction and temporal prediction were added to the proposed work. The PM concentration value is exceeding

the air quality identifies the warning alert for the pollution. The station B indicates less RMSE value than other stations. The performance measurement for the proposed system can be evaluated by Mean Absolute Error(MAE) and Root Mean Square Error(RMSE). List of stations with corresponding MAE and RMSE values are listed in Table 1. They are defined as follows

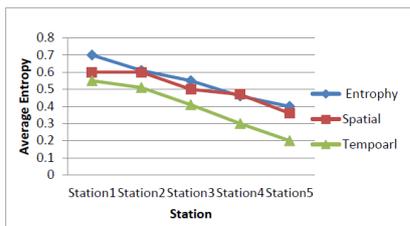
$$MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (3)$$

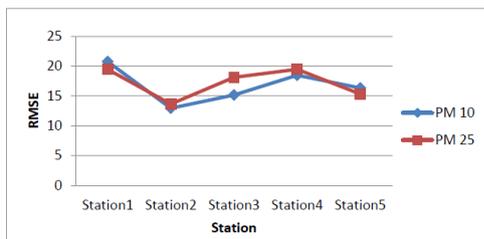
**Table 1 Parameters**

S.No	Station A	Station B	Station C	Station D	Station E
RMSE	20.75	12.94	15.16	18.46	16.34
MAE	10.98	12.45	14.59	14.6	16.7

We also calculated the predictive accuracy our proposed model with other approaches. Our approach achieves higher level accuracy than other methods. The entropy based search can be used for ranking the location. Inferred AQI values calculated at the particular interval. The locations will be based on smallest rank value. When we are applying temporal features, we select the locations with dissimilar meteorology features. Average entropy and RMSE analysis are described in fig 3&4. Pearson correlation can be used to measure the similarity between current location and existing stations. Spatial features also applied for our proposed approach.



**Fig 3 Average Entropy Analysis**



**Fig 4 RMSE Analysis**

## 5. CONCLUSION

In this paper, we have developed efficient machine learning method for air pollutant prediction. We have focused multi view hybrid model that combines the spatial and temporal features. The temporal classifier is based on a support vector machine and gaussian kernel function for estimating the current pollution information in the particular location. The spatial classifier is based on back propagation algorithm to model the spatial correlation between air

qualities of different stations. For future work, we will implement the test bed implementation.

## REFERENCES

1. Yunhao Liu, Xufei Mao, Yuan He, Kebin Liu, Wei Gong, and Jiliang Wang(2013),” CitySee: Not Only a Wireless Sensor Network”, IEEE Network Vol 27, pp.42-47.
2. Vijay Sivaraman, Blanca Gallego Luxan, and Ashfaqur Rahman” Design and Evaluation of a Metropolitan Air Pollution Sensing System, IEEE Sensors Journal, Vol. 16, NO. 5, 2016.
3. Khaled Bashir Shaban, Abdullah Kadri, and Eman Rezk LNCS ,” Urban Air Pollution Monitoring System With Forecasting Models”, IEEE Sensors Journal, VOL. 16, NO. 8,2016.
4. Gonçalo Marques and Rui Pitarma(2016),” An Indoor Monitoring System for Ambient Assisted Living Based on Internet of Things Architecture”, International Journal of Environment Research and Public Health, Vol, 13,pp.1-14.
5. Xiaoeng, QiLi Yajie, ZhuJunxiong, HouLingyan, JinJingjie (2015), Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, Vol.107, pp.118-128.
6. Klein M, Chang HH, Flanders WD, Strickland MJ(2014),” Classification and regression trees for epidemiologic research: an air pollution example”. Environ Health. Vol.13, pp.17-25.
7. Chen HW, Tsai CT, She CW, Lin YC, Chiang CF(2010),”. Exploring the background features of acidic and basic air pollutants around an industrial complex using data mining approach. Chemosphere”, Vol.81pp.1358–67.
8. Raja Vara Prasad Y , Mirza Sami Baig , Rahul K. Mishra(2011),” Real Time Wireless Air Pollution Monitoring System”, ICTACT Journal of Communication Technology, Vol.2,pp.370-375.