REAL TIME PREDICTION FOR FINE GRAINED AIR QUALITY MONITORING SYSTEM

UDAYA SHANKAR BUDAGAVI [1], SIRISHA K V [2], VISHAL M R [3] ,Prof SHEELA B.P[4], Dr SREEPATHI B[5]

[123] StudentS , Prof SHEELA B.P[4] ASSITANT PROFESSOR ISE DEPT, Dr SREEPATHI B[5] HOD,ISE DEPT,
**[12345] RAO BAHADUR Y MAHABALESHWARAPPA ENGINEERING COLLEGE, BALLARIDEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING*

ABSTRACT

Air pollution is one of the serious problems in the urban cities, where particulate matter (PM2.5) has greater effect on the humans than any other impure substances. The significant air pollution problem, monitoring and prediction for Air Ambiance have become increasingly necessary. To provide instantaneous close-grained Air Ambiance monitoring and prediction in urban areas, we have established our own Internet-of-Things-based sensing system in Peking University. Due to the energy constraint of the sensors, it is preferred that the sensors wake up alternatively in an asynchronous pattern, which leads to a sparse sensing dataset. In this paper, we propose a novel approach to predict the instantaneous close-grained Air Ambiance based on asynchronous sensing. The sparse dataset and the spatial-temporal meteorological relations are modeled into the correlation graph, in which way the prediction procedures are carefully designed. The advantage of the proposed solution over existing ones is evaluated over the dataset collected by our Air Ambiance monitoring system.

Keyword : -*Air Ambiance Monitoring, investigation, , Air pollution. Air Ambiance indexing, instantaneous, close-grained*

I. INTRODUCTION

Particulate matter is one of the dangerous parts of the air pollution. It is a composition of both liquid and solid particles; it may occur naturally due to volcanoes and fire burst or by human involvement such as power plants and vehicle emission. It occurs in various sizes; each size of PM has its own impact on humans as well as nature. It can easily penetrate into human blood, lungs and causes lot of diseases such as lung cancer and heart attack. Its presence in the air is causing lot of problems in the urban cities and it is increasing rapidly. Hence controlling or removing the particulate matter from air is the most important topics in the urban cities. Our study will provide crucial information of air pollution levels in a particular city by using AQI, so that it will be helpful for pollution control

In addition, air pollution modelling is used in science to help understand the relevant processes between emissions and concentrations, and understand the interaction. In this paper, we propose a novel scheme to conduct the close-grainedand instantaneous prediction of AQI based on asynchronous data collected by our monitoring system. By designing the correlation graph (CG), we present the asynchronous sensing data and the spatial-temporal-meteorological relations. Based on the CG model, the prediction procedures are carefully designed and an optimization problem arises. To obtain the instantaneous close-grainedprediction results, we aim to solve the optimization problem by an algorithm combining a closed-form derivation and genetic algorithm. The advantage of the proposed solution over existing ones is evaluated over the dataset collected by our monitoring system.

The main contributions of our work are listed as follows:

- We present our Air Ambiance system in Peking University based on massive IoT sensors, where asynchronous sensing data are modeled into a spatial-temporal-meteorological graph.

- We propose a novel prediction algorithm for instantaneous and close-grainedAir Ambiance based on sparse dataset.
- We evaluate the performance gain of our approach over existing methods based on real measured data

II. System model

2.1 Air Ambiance monitoring:

Air contains a mixture of gases, small solid and liquid particles. Some substances come from natural sources while others are caused by human activities. The air is said to be polluted when the contents of the air cause harm to the comfort or health of human and animals, or could even damage plants and other materials. These contents are termed as air pollutants and can be either particles, liquids or gaseous in nature. Air Ambiance Monitoring (AQM) is carried out to assess the extent of pollution, ensure compliance with national legislation, evaluate control options, and provide data for Air Ambiance modelling. The goal of AQM is to protect humans and the environment from harmful air pollution [5, 6]. There are different methods to assess any type of pollutant depending on the complexity, consistency and detail of data. These range from simple passive sampling techniques to highly sophisticated remote sensing devices.

The need for the implementation of AQM is to make mitigation strategies and arouse environmental awareness among citizens. Hence, several techniques and technologies have been introduced to monitor Air Ambiance . The increasing level of air pollution is mainly from sources such as smoke from vehicles exhaust and industrial activities. The common gases affecting the quality of air are carbon monoxide, sulphur dioxide, oxides of nitrogen, ozone, lead and dust particulates. Air Ambiance monitoring is therefore needed so that appropriate actions can be taken in order to mitigate its negative impact. Usually databases are used to store the collected data from a monitoring system. The data is then retrieved and analyzed to see if they are aligned to the pollution regulatory standards or not. In simple terms Air Ambiance monitoring network is used to record the concentration of pollutants and these information are delivered to the population to notify against danger.

Various technology and methodology have been used in order to provide Air Ambiance data in real time ranging from traditional way of passive sampling technique to the most sophisticated means such as use of sophisticated remote sensing devices. It is essential to define the options and monitoring methodology most appropriate in terms of cost, reliability and ease of operation. A means of monitoring air pollution is through online General Packet Radio Service (GPRS) sensors comprising of a microcontroller chip and an application server. The mobile data acquisition unit collects the pollution level and organise it into a frame with Global Positioning System (GPS) location, date and time. This frame is then uploaded to the GPRS modem and sent to the pollution server through the public mobile network .

Air Ambiance monitoring stations are often expensive and deliver a low resolution sensing data as these stations cannot be densely deployed. Alternatively, one of the effective solutions to provide real time pollution data is through the use of wireless sensor network (WSN) for Air Ambiance monitoring which is easy to set up and inexpensive. Consisting of calibrated sensors, WSN systems use a data aggregation algorithm and a routing protocol along with a light-weight middleware for transmission of the pollution data to a base station where they are visualized in graphical forms. Other parameters like humidity and temperature needs to be taken into consideration for providing more accurate pollutant data as these parameters affect the measured gas concentrations.

2.2 Air Ambiance indexing:

The assessment and calculation of air pollution level is built on standards which is present in almost every country of the world. The United States Environmental Protection Agency (EPA), the World Health Organization (WHO), the European Commission (EC).A suitable way for characterizing atmospheric pollution is through Air Ambiance index (AQI). AQI is a quantitative tool which provides information on how fresh or polluted the air is by consolidating the pollution data in the form of reports. Many countries make use of some type of AQI to interpret the quality of the air. An AQI is useful in several ways such as easy interpretation of Air Ambiance situation by the general public [10]. Moreover, based on the AQI quick actions can be undertaken, corrective pollution control strategies may be implemented from the trend of events, the impact of regulatory actions may be assessed and scientific researches may be carried out. The index values help to divide the air pollution situation into categories such that each category is identified by a simple informative descriptor which can be easily used to inform the public on the status of the air as shown in Table 1 .

Index values	Interval	AQI category
1	AQI > X_1	Very good
2	$X_2 < AQI < X_1$	Good
3	$X_3 < AQI < X_2$	Acceptable
4	$X_4 < AQI < X_3$	Poor
5	AQI < X_4	Bad

Table 1.
Summary of AQI range and descriptor .

III System Analysis

3.1 Existing System:

Air Ambiance Monitoring System Statistical models have been applied for air pollution prediction on the basis of meteorological data . However, existing studies on statistical modeling have mostly been restricted to simply utilizing standard classification or regression models, which have neglected the nature of the problem itself or ignored the correlation between sub-models in different time slots. On the other hand, machine learning approaches have been developing for over 60 years and have achieved tremendous success in a variety of areas . There exist various new tools and techniques invented by the machine learning community, which allow for more refined modeling of a specific problem. In particular, model regularization is a fundamental technique for improving the generalization performance of a predictive model.
Accordingly, many efficient optimization algorithms have been developed for solving various machine learning formulations with different regularizations.

3.2 Proposed System:

We paired the collected meteorological data and air pollutant data on the basis of time to obtain the required data format for applying the machine learning methods. In particular, for each variable, we formed one value for each hour. However, the original data may have contained multiple records or missing values at some hours. To preprocess the data, we calculated the hourly mean value of each numeric variable if there were multiple observed records within an hour and chose the category with the highest frequency per hour for each categorical variable if there were multiple values.

IV MODELING AND ANALYSIS

4.1 DATA DESIGN:

The dataset has been divided into a training set and testing set. Both the training set and the testing set contain Air Ambiance data without header. In this paper, we only focus on air monitoring data with the header. Due to the irrationality of the segmentation of the training set and the testing set in the original dataset, after merging the two datasets, the training validation set and the testing set are redivided. Users can have queries about the process. This part of project is dedicated to make and get response for queries that are needed to answerable. The major part of the modules is making project as interactive one, queries have been very normally arise to users regarding different details about the process. To make up for the above deficiency, it is recommended to deploy numerous low-cost tiny Internet-of-Things (IoT) sensing devices to monitor close-grainedair quality(mainly PM2.5 values) for the regions with complicated terrain.

4.2 ARCHITECTURAL DESIGN

Asynchronous Sensing System

We establish a wireless sensor network system for finegrained Air Ambiance monitoring. For power efficiency, the deployed IoT sensors are programmed to asynchronously monitor AQI (mainly PM2.5 values) until the batteries are dead. The IoT sensors have been deployed in the campus of Peking University since Feb. 2018. Each sensor uploads the AQI, the location and the time after monitoring . In addition, weather conditions are collected from the website of China Meteorological Administration . All features are preprocessed before being applied.

Real time prediction for fine grained air quality monitoring system architecture consist of following modules:

- **Initializing The Setup**
- **Starting the Application**
- **Dataset Input**
- **Feature Extraction**
- **Air Quality Analysis**
- **Accuracy Results**
- **Stop Analysis**

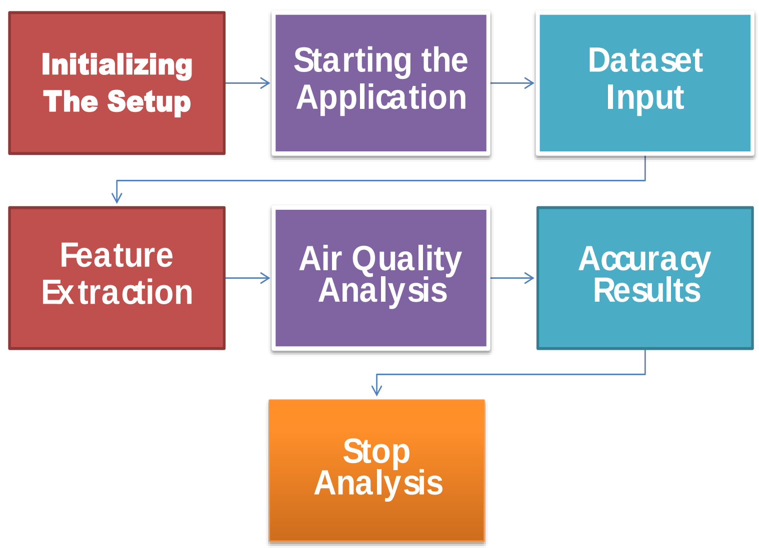


Fig 1 : *Block Diagram Of Fine Grained Air Quality Monitoring System*

4.3 Correlation Graph for Asynchronous

There are totally a number of M Air Ambiance sensors, denoted as {S1, · · · , Sm, · · · , SM}. Each sensor has a fixed 3D spatial coordinate (xm, ym, zm). For close-grainedconsideration, there are L POIs in concerned areas and we need to know the AQIs of the POIs, where M < L. Due to the limited number of the available sensors, the AQIs of the POIs without the sensors need to be estimated. Since the sensors monitor data asynchronously, we divide the system into short equal-length time slots, where each data can be considered as collected in a specific time slot. The collected data therefore could be very sparse, because only a small proportion of the sensors collect and upload data at any certain time slot.

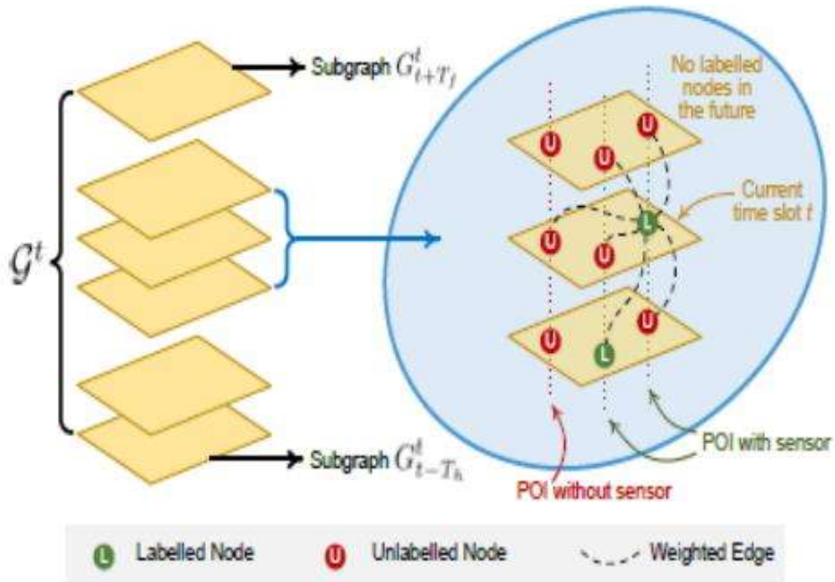


FIG 2: *: An illustration of the Correlation Graph*

After receiving the monitoring data at time slot t, the prediction system immediately executes
a round of iteration in order to obtain the instantaneous prediction F t , by using the history prediction F t-1 and the measured data {c t n} for the labeled nodes. This includes two procedures, namely the preparation procedure and the estimation procedure.

4.4 INTERFACE DESIGN :

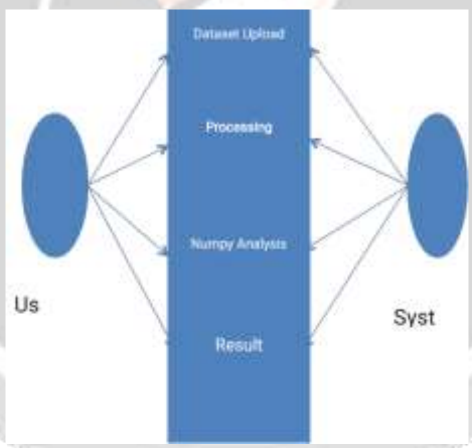


FIG 3:INTERFACE DESIGN OF AIR AMBIANCE SYSTEM

The above diagram gives the different modules how it is connected to the system perspective is given, as one can see all the modules are connected to the main admin system which will carry out all the functionality with easy.

V IMPLEMENTATION

The below table represents the implementation code the pseudo code includes code for loading the csv file datasets ,code for deleting all values which have null in type attribute,

```
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

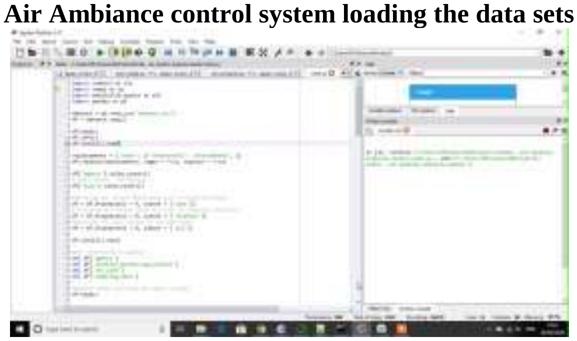
```
dataset = pd.read_csv('dataset.csv')
df = dataset.copy()
df.head()
df.info()
df.isnull().sum()

replacements = {'state': {r'Uttaranchal': 'Uttarakhand', }}
df.replace(replacements, regex = True, inplace = True)

df['agency'].value_counts()
# date format - mm/dd/yyyy
df['type'].value_counts()

#deleting all values which have null in type attribute
df = df.dropna(axis = 0, subset = ['type'])
# deleting all values which are null in location attribute
df = df.dropna(axis = 0, subset = ['location'])
#deleting all null values in so2 attribute
df = df.dropna(axis = 0, subset = ['so2'])
df.isnull().sum()
#not interested in agency
del df['agency']
del df['location_monitoring_station']
del df['stn_code']
del df['sampling_date']

#dataset after deleting the above columns
df.head()# 298 locations, 34 states
```

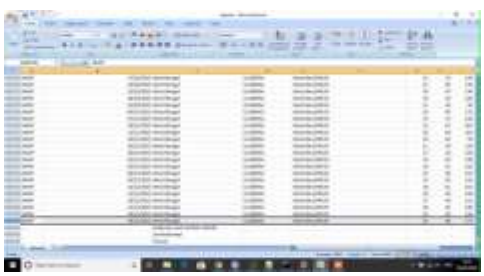
VI .CONCLUSION

In this paper, we have developed efficient machine learning methods for air pollutant prediction. We have formulated the problem as regularized MTL and employed advanced optimization algorithms for solving different formulations. We have focused on alleviating model complexity by reducing the number of model parameters and on improving the performance by using a structured regularizer. Our results show that the proposed light formulation achieves much better performance than the other two model formulations and that the regularization by enforcing prediction models for two consecutive hours to be close can also boost the performance of predictions. We have also shown that advanced optimization techniques are important for improving the convergence of optimizat on and that they speed up the training process for big data. For future work, we will further consider the commonalities between nearby meteorology stations and combine them in a MTL framework, which may provide a further boosting for the prediction.

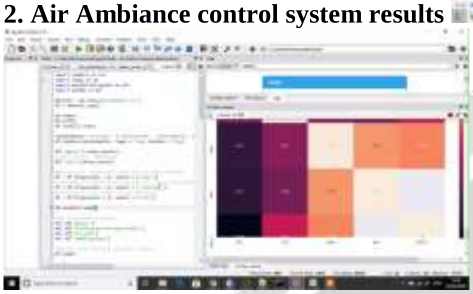
vANNEXURE

Air Ambiance control system loading the data sets





2. Air Ambiance control system results



ACKNOWLEDGEMENTS

The successful completion of my paper on " REAL TIME PREDICTION FOR FINE GRAINED AIR AMBIANCE MONITORING SYSTEM "would be incomplete without the mention of people who made it possible, we are extremely grateful to our respective Guide "SHEELA B.P" for her noble gesture, support co-ordination and valuable suggestions given to me in completing the paper. I also thank Dr. B.Sreepathi, H.O.D. Department of Information Science and Engineering, for extending all his valuable support and encouragement.

III. REFERENCES

1. Curtis, L.; Rea, W.; Smith-Willis, P.; Fenyves, E.; Pan, Y. Adverse health effects of outdoor air pollutants. Environ. Int. 2006, 32, 815–830. 2. Mayer, H. Air pollution incities.Atmos. Environ. 1999, 33, 4029–4037. 3. Samet, J.M.; Zeger, S.L.; Dominici, F.; Curriero,F.; Coursac, I.; Dockery, D.W.; Schwartz, J.; Zanobetti,
2. A. The national morbidity, mortality,and air pollution study. Part II: Morbidity and mortality from air pollution in the UnitedStates. Res. Rep. Health Eff. Inst. 2000, 94, 5–79. 4. Dockery, D.W.; Schwartz, J.; Spengler,
3. J.D. Air pollution and daily mortality: Associations with particulates and acid aerosols.Environ. Res. 1992, 59, 362–373. 5. Schwartz, J.; Dockery, D.W. Increased mortality inPhiladelphia associated with daily air pollution concentrations. Am. Rev. Respir. Dis. 1992,145, 600–604. 6. American Lung Association. State of the Air Report; ALA: New York, NY,USA, 2007; pp. 19–27. 7.
4. Environmental Protection Agency (EPA). Region 5: StateDesignations, as of September 18,2009.Availableonline:https://archive.epa.gov/ozonedesignations/web/html/region5desig.html (accessed on 17December 2017).