International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Multi Model Air Pollution Estimation for Environmental Planning Using Data Mining

A. Vinayagam[1], Dr. C. Kavitha[2], Dr. K. Thangadurai[3]

[1, 3] Assistant Professor, PG & Research Department of Computer Science,
Government Arts College (Autonomous), Karur - 639 005, Tamilnadu, India.

[2]Assistant Professor, Department of Computer Science, Thiruvalluvar Govt Arts College, Rasipuram, Tamil Nadu, India.

Abstract- ***Air pollution is an higher threat for not only the environmental condition of any country, also a biggest factor for global warming. In this decade , the nation planning commissions of every country shift their focust to get solutions for air pollution and trying to save the environment conditions of the country. We hope this is the first article which tries to provide some solution which could support environmental planning. We propose a multi model objective method which estimates the pollution could be surrounded according to many factors like population, geographic area, vehicle movements, factories , zonal temperature and etc. The proposed method computes the pollution range according to these factors and each has some weightage in computing the pollution range , because the lack of forest, more traffic, more factory outlets, population are the factors which highly affects the ratio of pollution occurring in air. The proposed method uses different other factors and computes the scarcity of forest and agri lands in particular region using their geographic area and other parameters. The propsed model identifies the most region which generates higher air pollution in a country and supports the planning commission to provide more schemes to increase the ratio of forest in that particular region.***

Keywords: Air Pollution, Global Warming, Data Mining, Pollution Estimation.

1. Introduction

The modern world is full of movements and the environmental conditions are getting changed at all the fractions of life. The people are moving to modern world and at most locations the forest are occupied by the human, they constructs building and resorts at the forest. This behavior reduces the forest region and produce lake of trees. The air pollution generated by the human being can be observed only through the forest where the concentration of pollution could be reduced. The modern people are not interested in cultivating trees and garden development. This makes trouble for the next generation and reduces the oxygen level in the air.

The hot topic at the environmental researchers is the global warming, the topic global warming is about the world temperature rise and how its developed and moving. The global temperature is increasing every year at 3 percentage and it caused by the osoan damage. The air pollution is the only factor for the ozoan damage and it caused by the human usage of cars, factories, traffics and etc.. The factories which are not reducing their carbon exhaust produces more pollution to the air and its increasing at daily basis. Also the lack of agri lands and trees also supports global warming and this makes difficult for the people to survive at the summar and they use airconditioners which also produces air pollution at low level.

The air pollution is estimated on various factors, simply saying , it could be estimated according to the vehicle movement on a region and how many vehicles are there in the region. Also the area or agri lands , forest, factories present in a geographic region. The air pollution control board has conducting survery in each year about the vehichle movement and factories of the region. The pollution value is estimated using all these factors and we propose a new model to estimate the pollution level using all these factors.

Data mining is a computing approach to extract information required from large set of data. In air pollution estimation, the data mining idealogy could be used where to mine information from large set of traffic data and to infer useful information about traffic pattern and how it affects the environment conditions. Similarly from other factors we could identify and extract useful informations from the large set of data.

2. Related Works

There are few works which has been proposed in this area and we discuss about them here. iMAP: Indirect measurement of air pollution with cellphones [1], introduce the cellphone-based indirect sensing problem. While participatory sensing aims at monitoring of a phenomenon by deploying a dense set of sensors carried by individuals, our indirect sensing problem aims at inferring the manifestations of a sparsely monitored phenomenon on the individuals. The main advantage of the indirect sensing method is that, by making use of existing exposure modeling and estimation methods, it provides a more feasible alternative to direct sensing. Collection of time-location logs using the cellphones plays a major role in our indirect sensing method, while direct sensing at the cellphones is unneeded.
SO2 classification for air quality levels estimation using artificial intelligent techniques[2], presents a new methodology to detect and classify SO_2 concentration according to the air quality level. In this classification, meteorological variables are analyzed to make a classification decision. The method consists of three steps. In first step, we group using a SOM neural networks the pollutant concentration in two classes, these classes are

noise data and validated data. In second step, we create a representative feature vector using the information contingency levels that we know a priori. In third step, a new SOM neural network is trained with the representative feature vector built in second step, then the pollutant concentrations and meteorological variables (validated data) are self-organized in fourth classes according to contingency levels. Finally, we obtained the air quality level.

Dynamic estimation of air pollution [3], propose advection-diffusion model of air pollution over an urban area. The region is subdivided into a grid and the three-dimensional partial differential equation of the pollution concentration is reduced to a linear vector difference equation. Along with this discrete equation, a stochastic model of air pollution is considered and pollution concentrations over the area are estimated from observed data generated by a few monitor points.

In Estimation of exhaust emissions of marine traffic using Automatic Identification System data [4], an Automatic Identification System (AIS) receiver is used to obtain ship data. AIS recognizes a vessel's Maritime Mobil Maritime Identify (MMSI), speed of ship, initial position of ship and ship type. This data is used to evaluate the marine traffic density in the Madura Strait area. Information from ship databases and AIS data are combined for retrieving gross tonnage (GT) information, which is then used to estimate the ship's air pollution emissions. Air pollution estimates also consider the ship's operation modes such as berthing, maneuvering and hotelling.

Air pollution data classification by SOM Neural Network [5], presents a Self-Organizing Maps (SOM) Neural Network application to classify pollution data and automatize the air pollution level determination for Sulphur Dioxide (SO2) in Salamanca. Meteorological parameters are well known to be important factors contributing to air quality estimation and prediction. In order to observe the behavior and clarify the influence of wind parameters on the SO2 concentrations a SOM Neural Network have been implemented along a year. The main advantages of the SOM is that it allows to integrate data from different sensors and provide readily interpretation results. Especially, it is powerful mapping and classification tool, which others information in an easier way and facilitates the task of establishing an order of priority between the distinguished groups of concentrations depending on their need for further research or remediation actions in subsequent management steps.
Influence of Air Pollution on Cardiovascular Diseases Prevalence in Developing Countries: An Eco-Social Model [6], proposes the use of eco-social analysis supported by Geographical Information System (GIS) to derive more accurate results. Local/Urban-Scale PM10 Concentration Estimation from TM Imagery [7], offers a unique opportunity to estimate air quality that is critically important for the management and surveillance of air quality in some cities of China, which have experienced elevated concentration of air pollution but lack adequate spatial-temporal coverage of air pollution monitoring. A local/urban-scale PM10 concentration estimation method is developed with 30m resolution TM imagery. The method can be used for local/urban-scale imagery under a variety of atmospheric and surface conditions.

Development of a real-time on-road emissions estimation and monitoring system [8], estimate and monitor operational on-road emissions with high accuracy and resolution in real time. The two sets of critical information for emission estimation, vehicle mix and vehicle activity, are directly generated from traffic detection using inductive vehicle signature technology. An initial implementation on a section of the I-405 freeway at Irvine, California is demonstrated. With more widespread deployment, the system can be used to perform before-and-after evaluation of certain mitigation strategies, to develop time sensitive optimal traffic control strategies with the purpose to control emissions, and to provide high fidelity greenhouse gas and air quality information to policymakers, researchers, and the general public.

All these approaches are focused only towards real time data and misses the earlier history to be visited before estimating the air pollution. We propose such a model to predict the upcoming air pollution with the help of time variant data and data mining techniques.

3. Proposed Method

An multi model air pollution estimation is an probabilistic model which comprise of three stages namely, preprocessing, pollution pattern generation, Pollution Estimation. The proposed method maintains time variant data about the pollution properties and based on the pollution information of pevious time zone current pollution has been estimated.

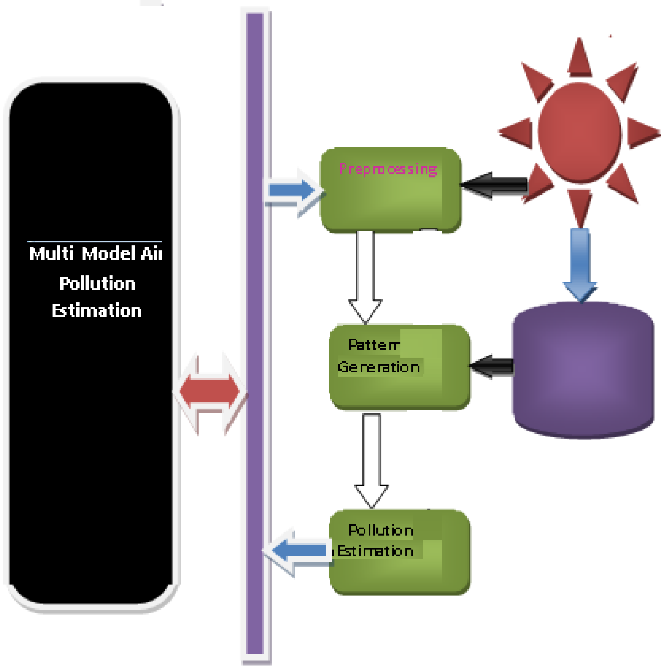


Figure 1: Proposed System Architecture

3.1 Preprocessing

At this stage , the time variant data which is stored at the history is retrieved and from the time variant data we split them into various time frames according to the precison we need. From each history of air pollution data , missing values are identified and cleaned or filled with substitute values. Finally the current details of multi attribute information about the air pollution is fetched to compute the estimation.

3.2 Pattern Geneneration

The pollution pattern is generated using the preprocessed information of pollution history. Each entry from each time frame of history is constructed as a pattern like time zone, Geographic region, total forest area, total agri area, total populations, total vehicles, fuel consumption, industries,zonal temperature. All these attributes are fetched from each of the history record and converted into form a vector. Now we have set of patterns , in which each represent a strategic information about a zone at particular time window. The time window may be of a week, month, quarterly, or annulay. Generated pattern is used to compute the probability of air pollution may occur at next time window.

Input: Preprocess Pollution History Ph.
Output:Pollution Pattern PoP.
for each record ph_i from Ph
extract the following information to construct the pattern or vector.
Pv = { zone, Geographic region, area of forest, area of agriculture, population, vehicles, fuel consumption, industries, zonal temperature}
PoP= Pv+ ΣPv(PoP).
end.

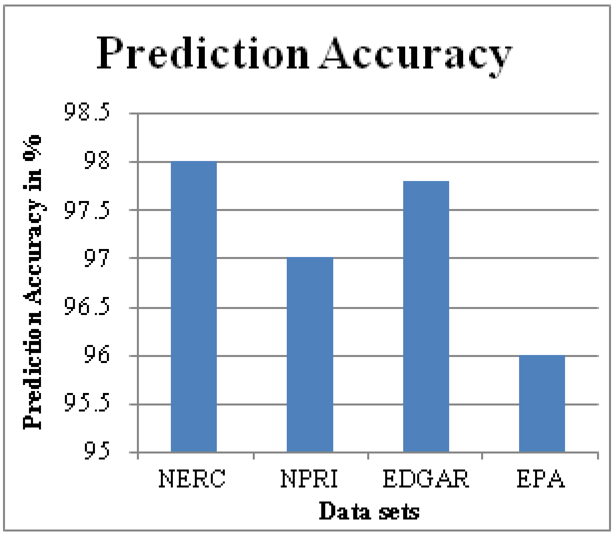
3.3 Pollution Estimation

The current pollution value and future pollution value is estimated in this by using pollution pattern generated in the previous step. The proposed method estimates the pollution using the probabilistic model. Because the pollution occurred in the previous time window has records but there is no record will be available for the current status and future happenings. So that the pollution will be estimated as follows.

Input: Pollution Pattern PoP.
Output: Current Pollution and Future Value.
Step1: Split time zone into N.
Step2: for each time zone
identify patterns $PoP_T = \int_{Tstart}^{Tend} PoP(p) \in p(Time)$
end.
Step3: for each time zone
compute average pollution occurrence Apc as follows.
compute geographic region factor Gr =($\sum_1^N PoP(Gr)$/N)×.($\sum_1^N PoP(Pollution)/N$.
Compute forest area factor Fr = ($\sum_1^N PoP(Fr)$/N)×.($\sum_1^N PoP(Pollution)/N$.
Compute Agriculture factor Ar = ($\sum_1^N PoP(Ar)$/N)×.($\sum_1^N PoP(Pollution)/N$.
Compute Popullation factor Pr = ($\sum_1^N PoP(po)$/N)×.($\sum_1^N PoP(Pollution)/N$.
Compute vehicle factor Vr = ($\sum_1^N PoP(Veh)$/N)×.($\sum_1^N PoP(Pollution)/N$.
Compute Fuel consumption factor Fcr= ($\sum_1^N PoP(Fc)$/N)×.($\sum_1^N PoP(Pollution)/N$.
Compute Industry factor Ir = ($\sum_1^N PoP(In)$/N)×.($\sum_1^N PoP(Pollution)/N$.
compute zone temperature factor Zr = ($\sum_1^N PoP(Zt)$/N)×.($\sum_1^N PoP(Pollution)/N$.
end.
Step4: construct computed values as a rule set.
Airpollution rule set Ars = Ars+ Ar(Gr,Fr,Ar,Pr,Vr,Fcr,Ir,Zr).
Step5: Read current details of record.
Identify each value of pollution record to which is belongs to.
for each record of time zone
identify between which value the attribute comes.
Select the appropriate factor value.
end.
Step6: Sum all the values of computed factors.
Current pollution Cp =($\sum Gr + Fr + Ar + Pr + Vr + Fcr + Ir + Zr)/8$
Step7: Future Pollution Ratio Fpr.
for each time zone
compute pollution ration Pratio = ($\sum Gr + Fr + Ar + Pr + Vr + Fcr + Ir + Zr)/8$
Add to Pollution matrix Pm(i) =Pratio.
end.
Step8: compute standard deviation of pollution.
Fpr = Cp× Ø(Pm).
Step9: Stop.

4. Results and Discussion

The proposed method has been evaluated with the data set available at. The proposed method has been tested for its efficieny in prediction and accuracy. We have used various size of data set to evaluate the performance of the proposed approach. We have used NERC data set , which is provided by open air project, Canada and NPRI data set of Canada.



Graph 1: shows the prediction accuracy achieved.

The graph1 shows the value of prediction accuracy achieved by the proposed method with different data sets. It is clear that the proposed method has produced higher efficient results with all the data sets available.

5. Conclusion

This paper presented an multi model approach for air pollution estimation using data mining. The pattern based data mining approach has been used for pollution estimation. The proposed method used all the factors of air pollution , so that to increase the performance of the pollution estimation. Unlike other methods we have used histories of air pollution and dynamic details also. The proposed method has produced efficient results.

References

[1] Demirbas M, iMAP: Indirect measurement of air pollution with cellphones, IEEE Conference on Pervasive computing and communications, 1-6, 2009.
[2] Cortina-Januchs, M.G, SO_2 classification for air quality levels estimation using artificial intelligent techniques, Multiconference on Electronics and photonics,pp:158-162, 2008.
[3] Desalu, Dynamic estimation of air pollution, IEEE Transaction on Automatic Control, vol:19,issue:6, pp:904-910,2003.
[4] Pitana,Estimation of exhaust emissions of marine traffic using Automatic Identification System data, IEEE OCEANS, 1-6, 2010.
[5] Barron-Adome,Air pollution data classification by SOM Neural Network, World Automation Congress, pp:1-5, 2012.
[6] Olayanju, L.O ,Influence of Air Pollution on Cardiovascular Diseases Prevalence in Developing Countries: An Eco-Social Model, Developments in E-System Engineering, pp:50-55, 2011.
[7] Weiwei Song ,Local/Urban-Scale PM10 Concentration Estimation from TM Imagery, IEEE conference on Bioinformatics and Biomedical Engineering (iCBBE), 2010.
[8] Hang Liu, Development of a real-time on-road emissions estimation and monitoring system, IEEE conference on intelligent transport system, pp:1821-1826, 2011.
[9] Sharma R.K.,Detection/ estimation of adulteration in gasoline and diesel using ultrasonics, International conference on industrial information system,pp:509-512,2007.
[10] Bingduo Yang ,Efficient local AADT estimation via SCAD variable selection based on regression models, International conference on control and decision, 2011.
[11] Yajie Ma, Air Pollution Monitoring and Mining Based on Sensor Grid in London, International conference on sensors, 2008.
[12] Richards, M.; Ghanem, M.; Osmond, M.; Guo, Y.; Hassard, H.. Grid-based Analysis of Air Pollution Data. *Ecological Modelling* 2006, *I94*, 274-286.