ISSN: 0975-766X
CODEN: IJPTFI
Available Online through
Research Article
www.ijptonline.com

AIR QUALITY MINING IN INDUSTRIAL AREA SUPPORT POLICY MAKERS

S.Christy*[1], Dr.V.Khanaa[2], Dr. Rajkumar Samuel[3]
[1]Research Scholar, Bharath Institute of Higher Education and Research, Bharath University, Chennai.
[2]Dean Information, Bharath Institute of Higher Education and Research, Bharath University, Chennai.
[3]Director, Hubert Enviro Care Systems Pvt Ltd, Chennai.
Email: christymelwyn@gmail.com

Received on: 15.10.2016 *Accepted on: 22.11.2016*

Abstract

Air pollution can affect our health and environment in many ways. In the past few years, the heavy environmental loading has led to the deterioration of air quality in industrial area near Chennai. The task of controlling and improving air quality has attracted a great deal of national attention. Ambient air quality data mining is a form of data mining concerned with finding hidden patterns inside largely available data, so that the information retrieved can be transformed into usable knowledge. This study aid data mining to uncover the hidden knowledge of air pollution distribution in the voluminous data retrieved from monitoring stations in an industrial area. The distribution of suspended particles like PM10, PM2.5, SO_2, CO polluted environment air are identified and could serve as an important reference for government agencies in evaluating present and devising future air pollution policies.

Keywords: Data mining; Air pollution monitoring; Air quality mining; Decision support.

1. Introduction

Data mining, also known as knowledge discovery in databases (KDD) is the process of discovering useful knowledge from large amount of data stored in databases, data warehouses, or other information repositories. Data mining, also known as knowledge discovery in databases (KDD) is the process of discovering useful knowledge from large amount of data stored in databases, data warehouses, or other information repositories[2]. Data understanding starts with an initial data collection and proceeds with activities to get familiar with the data, to identify data quality problems, and to discover first insights into the data.

Data preparation covers all activities that construct the final dataset to be modeled from the initial raw data. The tasks of this phase may include data cleaning for removing noise and inconsistent data, and data transformation for extracting the embedded features. The modeling phase applies various modeling techniques, determines the optimal values for parameters in models, and finds the one most suitable to meet the objectives.

The evaluation phase evaluates the model found in the last stage to confirm its validity to fit the problem requirements. No matter which areas data mining is applied to, most of the efforts are directed toward the data preparation phase [4]. In this study of mining air pollution data, our data preparation phase particularly emphasizes the data scale issue. The purpose of this study is to apply data mining technology to identify the national air quality distribution of Chennai, whose hourly air quality data are continuously collected and archived through a network of several stations. In dealing with voluminous data, we combine both wavelet transform (WT) and self-organization map (SOM) neural networks as our data mining technology[12].

The former is accredited with capability of investigating temporal variation with different scales, and the latter is known to be effective in isolating clusters in high-dimensional space. With both technologies, one can benefit from better understanding and interpretation of the pollution data.

2. Ambient Air Quality Data:

Ambient air quality data can be collected using instruments like Respirable dust sampler and Fine dust sampler and found various pollutants like PM10, PM2.5, SO_2, CO present around the industrial area. The samples are collected and calculated the amount of each pollutant in the air. Annual Arithmetic mean of minimum 104 measurements in a year at a particular industrial area taken twice a week 24 hourly at uniform intervals.

Whenever and wherever monitoring results on two consecutive days of monitoring exceed the limits for the respective category, it shall be considered adequate reason to institute regular or continuous monitoring and further several investigations. Data mining have been employed successfully to build a very important application in the field of air quality like predicting pollutants in the air.

3. Data Pre-processing

Data mining is an iterative process which involves various steps. A reference model for data mining will greatly help the success of the problem under investigation.

In this section, we present the process model used in mining air quality data by adapting the industry standard, the so called CRISP-DM model[14]. Fig. 1 shows the self-explanatory model and the associated steps[3]. An important step in the data mining process is data preprocessing. Prepare our data carefully to obtain accurate and correct results. First we choose the most related attributes to our mining task.

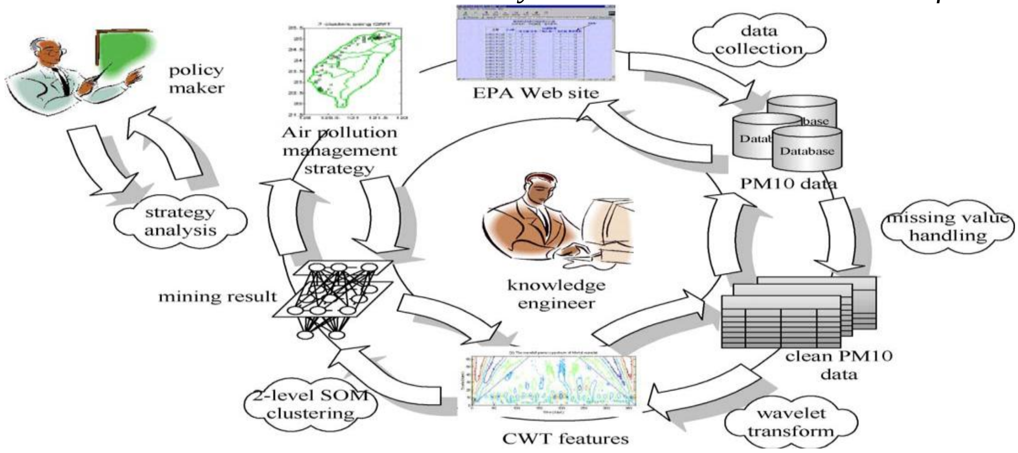

Fig. 1 The data mining process for air pollution.

Then we try to fill the missing with appropriate values. Because we are working with air data that is a form of time series, we must preserve the series smoothness and consistency. So we use linear interpolation method. This method is effective method to fill missing values in the case of time series where the missed value is strongly related to its previous and next values. After filling the missing values we apply windowing operation on monitored values in the case of classification and prediction**.**

4. Mining Significance

The enormous baseline data on ambient air quality is available and most of the times they are not utilized for the purpose they are intended. The application of mining will aid in the geographical clustering of the industrial establishments and also will support the policy makers to regulate the existing industries and lay new guidelines for the proposed and forth coming developments. The revised National Ambient Air Quality Standards 2009 supersedes the earlier notification in the number of parameters being monitored and also is more stringent on the earlier permissible limits. Table 1 shows the ambient air quality near Sriperumbudhur industrial area in the year 2012 (Average of 104 readings taken twice weekly).

Table-1: Ambient air analysis report.

S.No	Parameters	Units	Average value
1	PM_{10}	micro gm/m^3	55
2	$PM_{2.5}$	micro gm/m^3	28
3	Sulpur Dioxide	micro gm/m^3	20
4	Nitrogen Oxide	micro gm/m^3	25
5	Carbon Monoxide	mg/m3	<0.1

A repository of more than a decade ambient air quality data is available and will be used for the air quality mining which may be of support to the policy makers to take effective decisions.

5. Conclusion

The task of controlling and improving air quality has attracted a great deal of national attention. The government has since adopted an array of measures to combat this problem. This study will be taken forward with data retrieved from monitoring stations of the state of Tamil Nadu, India, which could serve as an important reference for the policy maker in formulating future policies.

References

1. Sarah N. Kohail, Alaa M. El-Halees, Implementation of Data Mining Techniques for Meteorological Data Analysis, International Journal of Information and Communication Technology Research, Volume 1 No. 3, July 2011.
2. Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27–34.
3. Li S., and Shue L., "Data mining to aid policy making in air pollution management," Expert Systems with Applications, vol. 27, pp. 331-340, 2004.
4. Pyle, D. (1999). Data preparation for data mining. Los Altos, CA: Morgan Kaufmann.
5. Rainsford, C. P., & Roddick, J. F. (1999). Database issues in knowledge discovery and data mining. Australian Journal of Information Systems, 6(2), 101–108.
6. Berkhin P., "Survey of clustering data mining techniques, Accrue Software, San Jose", CA, Tech. Rep., 2002.
7. Bilgin T., and Çamurcu Y., "A Data Mining Application on Air Temperature Database," Advances in Information Systems, Springer Berlin, Heidelberg, pp.68-76 .2004.
8. Central Intelligence Agency, "CIA World Factbook 2011," MobileReference. 2010.
9. Mangiameli, P., Chen, S., & West, D. (1996). A comparison of SOM neural network and hierarchical clustering methods. European Journal of Operational Research, 93(2), 402–417.
10. Martin-del-Brio, B., & Medrano, N. (1995). Feature map architectures for pattern recognition: Techniques for automatic region selection.
11. In D. W. Pearson, N. C. Steele, & R. F. Albrecht (Eds.), Artificial neural nets and genetic algorithms (pp. 124–127).

12. Mertins, A. (1999). Signal analysis-wavelets, filter banks, time-frequency transforms and applications. New York: Wiley.
13. Zhou, Z.-H. (2003). Three perspectives of data mining. Artificial Intelligence, 143(1), 139–146.
14. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard processmodel for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29–39.
15. Zhou, Z.-H. (2003). Three perspectives of data mining. Artificial Intelligence, 143(1), 139–146.
16. Mitra, S., Pal, S. K., & Mitra, P. (2002). Data mining in soft computing framework: A survey. IEEE Transactions on Neural Networks, 13(1), 3–14.

Corresponding Author:

S.Christy*[1],

Email: *christymelwyn@gmail.com*