International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-4 Issue-6 December-2012

Air Pollution Monitoring & Tracking System Using Mobile Sensors and Analysis of Data Using Data Mining

Umesh M. Lanjewar[1], J. J. Shah[2]
Department of Computer Science & Engineering[1]
Department of Information Technology[2]
G. H. Raisoni College of Engineering, Nagpur, India[1,2]

Abstract

This study proposes air pollution monitoring system and analysis of pollution data using association rule data mining technique. Association rule data mining technique aims at finding association patterns among various parameters. In this paper, association rule mining is presented for finding association patterns among various air pollutants. For this, Apriori algorithm of association rule data mining is used. Apriori is characterized as a level-by-level complete search algorithm. This algorithm is applied on data captured by various gas sensors for CO, NO2 and SO2 sensors. As association rule mining can produce several sequence rules of contaminants, the proposed system design can enhance the reproducibility, reliability and selectivity of air pollution sensor output.

Keywords

Air Pollution Monitoring, Association rule data mining technique, Apriori algorithm.

1. Introduction

Today's one of the major public health and environmental concern is air pollution. Air pollution can affect many body organs and systems in addition to the environment. According to the report of the World Health Organization (WHO), air pollution is a significant risk factor for multiple health conditions including skin and eye infections, irritation of the nose, throat and eyes. It also causes serious conditions like heart disease, lung cancer, pneumonia, bronchitis, difficulty in breathing and coughing due to aggravated asthma [1].

Air pollution is also the major cause for many premature deaths. The World Health Organization concludes that 2.4 million people die each year from causes due to air pollution. Air pollution not only has worse effect on people's health but also on the environment and can lead to acid rain, smog, deterioration of the ozone layer and global warming. So, it becomes very essential to monitor and control the air pollution. The best way to control air pollution is to monitor exceeding levels of air pollutants and by taking appropriate actions to control it. Several techniques can be used to monitor air pollution data, one such technique is Data mining.

Data mining refers [9] to the mining or discovery of new information in terms of patterns or rules from vast amounts of data. It involves the use of sophisticated data analysis tools to discover previously unknown patterns and relationships in large amount data. These tools can include statistical models, various mathematical algorithms, and machine learning rules. It involves algorithms that improve their performance automatically through experience. Data mining is not only collecting and managing data but also includes analysis and prediction.

2. Literature Review

Many air pollution systems in various areas were reported in recent literatures.

In 2011, Dan Stefan Tudose, Traian Alexandru Patrascu, Andrei Voinescu, Razvan Tataroiu, Nicolae Tapus et al. [1] proposed an environmental air pollution monitoring system that measures CO2, NO2, CO, HC & NH4 concentration using mobile sensors in urban environment. The acquired information about air pollution in surroundings is then stored on central on-line repository system periodically. It uses a wireless GSM modem connection for transferring data to a central computer. Also, the application can share the data publicly by displaying it on a dedicated web site.

In 2011, Diego Mendez, Alfredo J. Perez, Miguel A. Labrador, Juan Jose Marron et al. [3] proposed a participatory sensing system for air pollution monitoring and control using cellular phones, GPS technology and sensors to form a bidirectional mobile sensing information system. It mainly uses many

cellular phones to acquire large amounts of data in a simple and cost effective manner.

In 2012, Amnesh Goel, Sukanya Ray, Prateek Agrawal, Nidhi Chandra et al. [4] proposed a wireless sensor network to monitor air pollution levels of various pollutants due to environment changes. A wireless network is comprises of large number of sensors modes. This system proposes a method which mainly focuses on longer sustain time period of sensor network by effectively managing energy in sensor network, effectively processing of collected information and less overhead in transferring information between various sensor nodes.

In 2011, Wenhu Wang, Yifeng Yuan, Zhihao Ling et al. [5], in order to comply with requirements of oil and gas industry, an air quality monitoring system was proposed based on ZigBee wireless sensing technology. It uses ZigBee wireless network to send results to the monitoring center so that, if some abnormal situations happens, a quick warning will be generated to remind staff to take effective measures to prevent major accidents and protect human lives in industry.

In 2009, Ebrahim Sahafizadeh, Esmail Ahmadi et al. [6], proposed a system with k- means clustering and 5 clusters to cluster 53 year of cluster data from 1951 to 2003 of Air Pressure, Air humidity and dusty days per month. For this purpose thy have used Clementine software. Dusty days are classified into 5 classes and decision rule has been exported between air pressure, air humidity and dusty days of January, February and March of each year and other month of year dusty days.

Some of the above mentioned air pollution and quality monitoring systems based on sensors that reports pollutants levels are already exist and these are publicly available. But these systems are not mobile systems and perform measurements at fixed locations. Thus they do not cover the entire area of a city. Some of these systems uses wired modem, router, or short range wireless access point to transfer data to pollutant server resulting in bulky and costly systems.

3. Air Pollution Monitoring

Air pollution leads to instability, harmful and undesirable effects in the environment [4]. With the rapid growth of industrialization, environmental pollution has become a large area of concern. The primary pollutants are:

- Carbon Monoxide (CO): Carbon monoxide is a very poisonous gas. It is produced by incomplete combustion of fuel such as coal or wood, natural gas [4]. Vehicular exhaust is one of the primary sources of carbon monoxide. Road vehicles produce 91% of all CO emissions. When inhaled it can reduce the oxygen carrying capacity of one's blood and can cause headache or fatigue.
- Nitrogen Oxides (NO_x): It is a toxic gas and is one of the primary air pollutants. Especially Nitrogen-dioxide which are emitted from high temperature combustion in industry as well as from vehicles [4]. NO_x emissions are responsible for 'acid rain'. They also combine with hydrocarbons to form low level ozone and may cause lung disease.
- Sulfur Oxides (SO_x): Especially Sulfur-dioxide (SO_2) which is produced in various industrial processes. The oxidation of SO_2 in the presence of catalysts like NO_2 forms H_2SO_4, which is acid rain and causes damage to environment [4].

For the control of the air pollution in the environment it needs continuous monitoring of the quantity of gases present in the environment. This helps government officials, tourist and insurance companies, international organizations, and individuals to access the pollution data [2].

4. Data mining techniques

Data mining has been defined [7] as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data". It employs various computer techniques such as supervised or unsupervised learning algorithm techniques, in order to automatically search large data and derive patterns that can be used for either predictive (classification/regression) or descriptive tasks (clustering, association rule mining, etc.).

Data mining can be performed on data represented in quantitative, textual, or multimedia forms [6]. Data mining applications can use a variety of parameters to examine the data. They may include various association patterns where one event is connected to another event, such as purchasing a tooth paste and purchasing tooth brush, sequence or path analysis

(patterns where one event leads to another event, such as coming of festive sessions and purchasing of cloths), classification (identification of new patterns), forecasting (discovering patterns from which one can make reasonable predictions regarding future activities), and clustering (finding and visually documenting groups of previously unknown facts).

5. Association Rule Mining

The association rule mining is a very important and also the most active branch of data mining. Association rule refers to the rules of certain association relationship between groups of objects in the database.

A. Association rule concept:

Association rule mining can be described in [9] as following: assuming {i1,i2,i3,…,in} is n aggregates with different terms, then for a transaction database D, each element T in D is a set composed by some terms in I , $T \subseteq I$. The association rule is expressed as $X \Rightarrow Y$, here, $X \subset I$, $Y \subset I$, and $X \cap Y = \Phi$. The association rule mining is to discover all condition implicative expression meeting the minimum degree of confidence and support.

The confidence and support degree of these rules are all greater than or equal to the minimum degree of confidence and support. The confidence and support degree of the association rules respectively reflects the correct degree and the support rate of them.

In general, the user can define two thresholds which are respectively set as minimum confidence threshold and minimum support threshold, the confidence and support degree the data mining system generated are required to be not less than the two given thresholds, and then we can say that this rule is valid, otherwise it is not void or null. Thus a specific association rules can be uniquely identified by using an implication expression and two thresholds.

B. Association rule mining steps:

The steps of the mining association rules can be roughly described by a two-step process [9].

(1) Identify all the frequent term sets. That is, to identify all the term sets whose support degree greater than pre-given support threshold.
(2) To generate strong association rules on the basis of the found frequent term sets. That is, to generate those association rules whose support and confidence respectively greater than or equal to the pre-given support threshold and confidence threshold.

In the above two steps, it only needs to list all possible association rules on the basis of the found frequent term sets, and then measure these association rules by using support and confidence threshold, at the same time the association rules meeting the support and confidence threshold requirement are considered interesting association rules.

6. Apriori algorithm

Apriori is an algorithm to find all sets of items (itemsets) that have support greater than minimum support. The support for an itemset is defined as the ratio of the number of transactions that contain the itemset to the total number of transactions. Itemsets that satisfy minimum support constraint are called frequent itemsets. Apriori is characterized as a level-by-level complete search algorithm (breadth first search).

The algorithm makes multiple passes over the data. In each subsequent pass, a seed set of itemsets found to be frequent in the previous pass is used for generating new potentially frequent itemsets, called candidate itemsets, and their actual support is counted during the pass over the data. At the end of the pass, the itemsets those satisfying minimum support constraints are collected. These are the frequent itemsets and they become the seed for the next pass. This process is repeated again and again until no new frequent itemsets are found. The number of items in an itemset is called its size and an itemset of size k is called a k-itemset.

In Apriori, it first of all, find out the frequent term set with length of 1 which is recorded as l1, this l1 is used to find the aggregate l2 of frequent 2-term sets, then this l2 is used to find the aggregate l3 of frequent 3-term sets, and so on the cycle continues, until no new frequent k - term sets can be found. Finding each lk needs a database scan. Afterward, according to the minimum confidence threshold the effective association rules can be constructed from the frequent term sets [9]. Apriori algorithm flowchart is as shown in Figure 1.

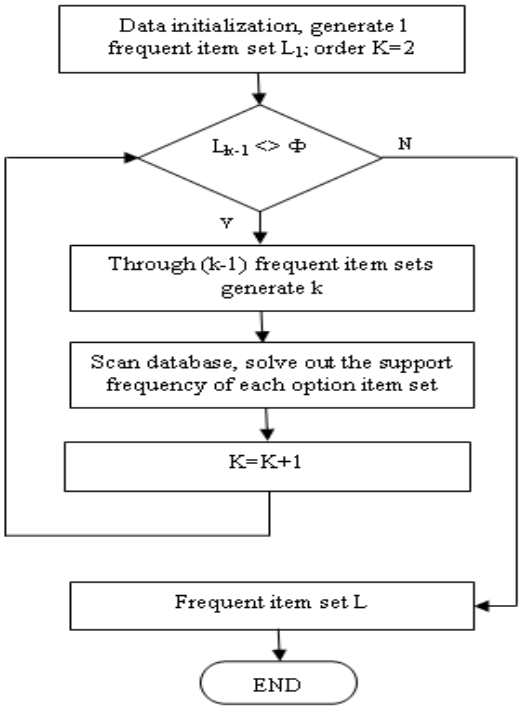

Figure 1: Apriori algorithm flowchart

7. Conclusion

This paper presents air pollution monitoring system using various gas sensors and tracking of the polluted area using global positioning system module. The system uses public transport buses to acquire air pollutants levels in large city area.

Many people will immediately benefit from this system. They may include asthmatic patients, joggers, or people concerned about the air quality. The government agencies that regulate and impose pollution standards can benefit from the large amount of analyzed data which may result in better understanding of the many pollutants affects the urban environment. It is also helpful to air quality management and to pinpoint major polluting sources at various places inside of a city. By the application of association rule data mining technique, it becomes possible to associate various air pollution data with each other. Thereby, one can predict air pollution patterns and to know the frequency of air pollutants exceeding standard levels in a particular area of a city.

Further, more complex logic could be incorporated in the application allowing the automated identification of problem area in the city.

References

[1] Dan Stefan Tudose, Traian Alexandru Patrascu, Andrei Voinescu, Razvan Tataroiu, Nicolae Tapus, "Mobile Sensors in Air Pollution Measurement", 2011 8th Workshop on Positioning Navigation and Comm., pp. 166-170, Apr. 2011.

[2] A.R. Al-Ali, Imran Zualkernan, and Fadi Aloul, "A Mobile GPRS-Sensors Array for Air Pollution Monitoring", IEEE Sensors Journal, vol. 10, no. 10, pp. 1666-1671, Oct. 2010.

[3] Diego Mendez, Alfredo J. Perez, Miguel A. Labrador, Juan Jose Marron, "P-Sense: A Participatory Sensing System for Air Pollution Monitoring and Control", IEEE International Conference on PERCOM Workshops, pp. 344-347, Mar. 2011.

[4] Amnesh Goel, Sukanya Ray, Prateek Agrawal, Nidhi Chandra, "Air Pollution Detection Based On Head Selection Clustering and Average Method from Wireless Sensor Network", 2012 Second International Conference on Advanced Computing & Communication Technologies, pp. 434-438, Jan. 2012.

[5] Wenhu Wang, Yifeng Yuan, Zhihao Ling, "The Research and Implement of Air Quality Monitoring System Based on ZigBee", 2011 7th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1-4, Sept. 2011.

[6] Ebrahim Sahafizadeh, Esmail Ahmadi, "Prediction of Air Pollution of Boushehr City Using Data Mining", 2009 Second International Conference on Environmental and Computer Science, pp. 33-36, Dec. 2009.

[7] Fani A. Tzima, Kostas D. Karatzas, Pericles A. Mitkas, Stavros Karathanasis, "Using data-mining techniques for PM10 forecasting in the metropolitan area of Thessaloniki, Greece", Proceedings of International Joint Conference on Neural Networks, pp. 2752-2757, Aug. 2007.

[8] Azuraliza Abu Bakar, Nurfathehah Idris, Abdul Razak Hamdan, Zalinda Othman, Mohd Zakri Ahmad Nazari,Suhaila Zainudin, "Classification Models for Outbreak Detection in Oil and Gas Pollution Area", 2011 International Conference on Electrical Engineering and Informatics, pp. 1-6, July 2011.

[9] Liu Hong-min, "Study and Implementation of Association Rule Algorithm in Data Mining", 2009 International Conference on Signal Processing Systems, pp. 821-825, May 2009.

Umesh M. Lanjewar received the B. E. degree in Electronics Engineering from Nagpur University, Nagpur, India. He is currently the student of Master of Engineering in Embedded Systems and Computing under the department of Computer Science and Engineering from G. H. Raisoni college of Engineering, Nagpur, India. He has industrial exposure and worked as software trainer. His research interest are embedded sytems, intelligent data acquisition techniques, wireless remote monitoring and control using GSM, GPRS networks.

J. J. Shah received her the B. E. degree in Computer Science and Engineering from Amravati University, Amravati, India, and the M. E. degree in Embedded Systems and Computing under the department of Computer Science and Engineering from G. H. Raisoni college of Engineering, Nagpur, India. She is currently pursuing Ph. D. degree in Computer Science and Engineering.

She is a Assistant Professor with Department of Information Technology in G. H. Raisoni College of Engineering, Nagpur, India. She has 8 publications in international journals and conferences. Her research interest are in the area of Data mining and soft computing, Embedded Sytems, Cloud Forensics. She is a IEEE member.