International Journal of Engineering& Scientific Research
Vol.8 Issue 07, July 2020,
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

FORECAST OF URBAN AIR POLLUTION USING MACHINE LEARNING ALGORITHM

NAGARAJU T A
Assistant Professor
Dept.of Electronics and Communication Engineering
Government Engineering College
Ramanagara, India

MANJUNATH H R
Assistant Professor
Dept.of Information Science and Engineering
Alva's Institute Of Engineering and Technology
Mangaluru,India

***Abstract*— Forecasts of daily pollutant levels have become a standard part of weather predictions in television, on-line, and in newspapers. Research groups also need to analyses larger timeframes across more locations to correlate long term developments for different pollutants with multiple serious health effects such as asthma. A system for monitoring and forecasting urban air pollution is presented in this paper. Data mining is the discovery of interesting, unexpected or valuable structures in large datasets. Experiments are implemented with different features groups and collaborative filtering algorithm in machine learning method for few cities in Karnataka. The focus of this paper is on the monitoring system and its forecasting module**

***Keywords*— Air Pollution Prediction, Machine learning Algorithm**

I. INTRODUCTION

It is widely believed that urban air pollution has a direct impact on human health especially in developing and industrial countries, where air quality measures are not available or minimally implemented or enforced [1]. Recent studies have shown substantial evidences that exposure to atmospheric pollutants has strong links to adverse diseases including asthma and lung inflammation [2].Considering the significance of air quality on human lives, the World Health Organization (WHO) has developed guidelines for reducing the health effects of air pollution on public health by setting the limits of the concentrations of various air pollutants, some of which are ground–level ozone (O3), nitrogen dioxide (NO2), and sulphur dioxide (SO2).

Traditionally, the concentrations of air pollutants are measured using air quality monitoring (AQM) stations that are highly reliable, precise, accurate, and are able to measure a wide spectrum of pollutants using standardized analysers. The main contribution of this paper is to present a prediction system for the air pollutants in the atmosphere using collaborative filtering algorithm for air quality simulations.

The data set used in this analysis is air pollution data collected from a government website across the states of India and for a variety of pollutants at regular time intervals. The Air Quality

Index is focused in this paper. The Air Quality Index (AQI) is a quantitative method to profile air pollution level. The daily AQI is an index for reporting daily air quality [2]. It is

determined by the maximum value of the Individual Air Quality Index (IAQI), which is calculated from mass concentrations of PM 2.5, SO2, NO2, CO and O3 in ambient air respectively; furthermore, the pollutant with maximum IAQI is called the primary pollutant. AQI is measured at monitoring stations throughout cities and reported daily by the government website. The remainder of this paper is organized as follows: Section II introduce several important air pollutants forecasting methods. Section III describes the proposed evaluation framework. Section IV gives numerical results. Finally, Section V provides conclusions.

II. BACKGROUND

A. Machine Learning

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data. The process of machine learning is similar to that of data mining [7]. Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new but one that's gaining fresh momentum.

B. Machine Learning and its applications

There are numerous applications of machine learning. It's actually hard to realize how much machine learning has achieved in real world applications. Machine learning is generally just a way of fine tuning a system with tunable parameters [7]. It is a way of making a system better with examples, usually in a supervised or unsupervised manner.

Neighborhood models are heuristics based models which uses similarity metrics, for e.g.: Pearson similarity, cosine similarity, for finding similar users and items. It is based on, very reasonable, heuristic that a person will like the items that are similar to previously liked items. Rating prediction in item based neighborhood models is given by weighted average of ratings on similar items [5].

C. Air pollutants measurement

Recent development of electronics has realized the vision of using wireless communication in devices used for monitoring wide range of real life parameters, such as temperature, pressure, and air pollution. These devices send their measurements wirelessly to a database hosted on a remote server for further processing and analysis [3]. The concept of using small size, inexpensive AQM motes that wirelessly communicate their air pollution measurements has been widely studied and implemented. The trend is moving towards the employment of the Next Generation of Air Monitoring (NGAM) that has the potential to complement the traditional AQM stations with small sized and inexpensive AQM motes that incorporate an array of gaseous sensors [4]. This paper focuses on presenting a pilot NGAM, and on the development of accurate forecasting models for predicting future average concentrations of some urban air pollutants, namely: O3, NO2, and SO2 all of which are mentioned as being harmful in the WHO's guidelines. The concentration of the air pollutants is obtained using an open source government website, http://www.cpcb.gov.in/CAAQM/mapPage/frmindiamap.aspx.

III. ALGORITHM

Collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. [6]. Applications of collaborative filtering typically involve very large data sets. Collaborative filtering methods have been applied to many different kinds of data including: sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources or in electronic commerce and web applications where the focus is on user data, etc.

All locations are weighted with respect to similarity with the prediction location. Similarity between locations is measured as the Pearson correlation between ratings vectors. Select n locations that have the highest similarity. Compute a prediction, Pa,u from a weighted combination. Similarity between two locations is computed using the Pearson.

$$P_{a,u} = \frac{\sum_{i=1}^{m} (r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^{m} (r_{a,i} - \bar{r}_a)^3 \times \sum_{i=1}^{m} (r_{u,i} - \bar{r}_u)^3}} \quad (1)$$

Where ra,i is the measurement given to item i by location a; and ra is the mean rating given by location a. The predictions are computed as the weighted average of deviations from the neighbour's mean: Where Pa,i is the prediction for the location a for item i. Pa,u is the similarity between location a and u. n is the number of location in the neighbourhood

.

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^{n} (r_{u,i} - \bar{r}_u) \times P_{a,u}}{\sum_{u=1}^{n} P_{a,u}} \quad (2)$$

IV. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system.

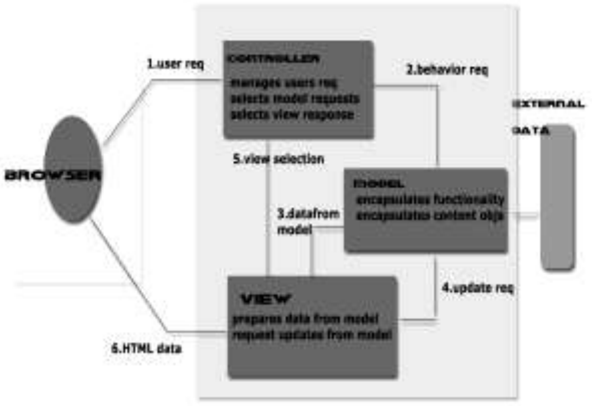

Fig 1: A J2EE uses MVC architecture.

MVC stands for Model View and Controller. It is a design pattern that separates the business logic, presentation logic and data. Controller acts as an interface between View and Model. Controller intercepts all the incoming requests. Model represents the state of the application i.e. data. It can also have business logic. View represents the presentation i.e. UI (User Interface). The proposed system architecture is based on the MVC architecture. In the proposed system, work addresses the question of how to predict fine particulate matter given a combination of weather conditions.

Therefore, this study aims to elaborate a statistical model to predict the pollution levels from the meteorological conditions. This statistical approach is based on data mining, which is searching for some patterns in raw big data in order to extract regularities that can be used to build a predictive model. In the proposed system by implementing the Collaborative filtering algorithm, mining all the previous data about the pollution details and predicting the next day's fine particulate matter of the environment.

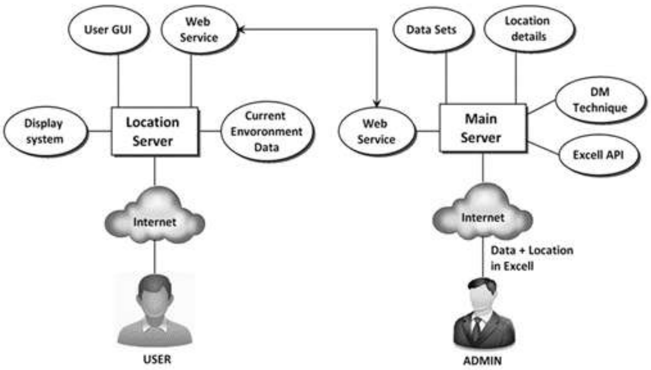

Fig 2: Architecture of System

V. PROGRAMMING APPROACH

A. Tools used

The IDE that is being used for the simulation of the project is Eclipse Mars 4.5. **Eclipse** is an integrated development environment (IDE) used in computer programming, and is the most widely used Java IDE. It contains a base workspace and an extensible plug-in system for customizing the environment. Eclipse is written mostly in Java and its primary use is for developing Java applications [7].

Apache Tomcat, often referred to as **Tomcat Server**, is an open-source Java Servlet Container developed by the Apache Software foundation (ASF). Tomcat implements several Java EE specifications including Java Servlet, Java Server Pages (JSP), Java EL, and Web Socket, and provides a "pure Java" HTTP web server environment in which Java code can run.

SQLyog is a GUI tool for the RDBMS MySQL. SQLyog is distributed both as free software free of charge as well as several paid, proprietary, versions. It is programmed and developed in C++ using Win32 API, no dependencies on runtime. It uses MySQL C API to communicate with MySQL servers, no dependencies on 'database abstraction layers' (like ODBC/JDBC).

B. Technologies used

Java Server Pages (**JSP**) is known as a technology that helps software developers create dynamically generated web pages based on HTML, XML, or other document types. To deploy and run Java Server Pages, a compatible web server with a servlet container, such as Apache Tomcat or Jetty, is required. A Java servlet processes or stores a Java class in Java EE that conforms to the Java Servlet API, a standard for implementing Java classes that respond to requests. Servlets could in principle communicate over any client–server protocol, but they are most often used with the HTTP protocol. Servlets can be generated automatically from Java Server Pages (JSP) by the Java Server Pages compiler. The languages that are being implemented are Java and SQL. Java is being used for servlets and JSP whereas SQL is used for JDBC connection between SQLyog and Eclipse.

Admin Module: In this module the admin manually uploads the city details i.e he can add or delete a city, admin has a info page, login page, adds city details i.e the present pollutant values for the particular day. Here the admin manually takes the data from Central Pollution Control Board (CPCB) and uses a excel sheet to upload the data into the main server, so that this data will used by the prediction algorithm to predict the future days data, which will be sent to the user when the user asks for prediction in the future

User Module: The user side module is available for all its users. The user has his login page which will have a username and password which he earlier used while registration. The user can enter the city details of which page he wants to get details of. This request is sent to the main server through web services. The request is processed and the prediction is done and results are sent back to the user and results will be displayed.

Prediction Process: When the user requests for predicted values, the request is sent to the main server. In the main server the dataset uploaded by the user is taken and Collaborative filtering is used to get the values. This resulting values are sent back to the user.

VI. CONCLUSION

Air quality is an important problem that directly affects human health. Air quality data are collected wirelessly from monitoring motes that are equipped with an array of gaseous and meteorological sensors. These data are analyzed and used in forecasting concentration values of pollutants using intelligent machine to machine platform. The platform uses collaborative filtering algorithm to build the Forecasting models by learning from the collected data. These models predict the concentration values. Due to the fact that the assessment of air pollution in cities is a fundamental problem in terms of public health, this study proposed to look for a model to predict the concentration of fine particulate matter (PM2.5) from meteorological data in the metropolitan area. Experimental results indicate that the more feature used, the more possibility to enhance the accuracy.

REFERENCES

[1] [1] Environmental Protection Agency. Health effects of ozone in the general population, 2015. http://www3.epa.gov/apti/ozonehealth/population.html.

[2] [2] C.A. Pope, C. A., and D.W. Dockery, "Health effects of fine particulate air pollution: lines that connect," Journal of the Air & Waste Management Association, vol. 56(6), pp. 709-742, 2006.

[3] [3] L. Draschkowitz, C. Draschkowitz, and H. Hlavacs, "Using video analysis and machine learning for predicting shot success in table tennis," EAI Endorsed Transactions on Creative Technologies, vol. 2, e2, 2015.

[4] [4] R. Holte, "Very simple classification rules perform well on most commonly used datasets," Machine Learning, vol. 11, pp. 63-91, 1993.

[5] [5] I.H. Witten, E. Frank, and M.A. Hall, Data Mining - Practical Machine Learning Tools and Technologies, 3rd ed. Burlington, MA: Morgan Kaufmann Publishers, 2011.

[6] J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000.

[7] I. H. Witten, E. Frank, M. A. Hall, and G. Holmes, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. San Mateo, CA, USA.

[8] Usha Mahalingam, Kirthiga Elangovan, Himanshu Dobhal, Chocko Valliappa, Sindhu Shrestha5, and Giriprasad Kedam "A Machine Learning Model for Air Quality Prediction for Smart Cities", IEEE 2019

[9] Mansi Yadav, Suruchi Jain and K. R. Seeja, "Prediction of Air Quality Using Time Series Data Mining", Springer 2019.

[10] Nidhi Sharmaa, Shweta Tanejab, Vaishali Sagarc, Arshita Bhattd, "Forecasting air pollution load in Delhi using data analysis tools.", Elseviere ICCIDS 2018.