International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Air Quality Sensing Network in Smart City using Lora Wan

Shriram S., P. Sai Prasanth, Meghna Anand

***Abstract:** Nowadays urban residents, especially ones in the city are more aware of air quality and how it can affect their health. Due to the limitation of spare sense nodes, the air quality information is very coarse in resolution that is not greatly understood nor accepted by the community. The demand for better resolution, information and prediction of air quality in these areas has brought us to this solution. Our proposed system gives us a real-time and fine-grained data map of the air quality in the city by using crowdsourced automobiles as well as their built-in sensors, which significantly improves the feasibility and practicability of the system in a more dynamic way. By using LoRa WAN network, we can increase the range without worrying about the amount of sensor data to be transmitted. Once the data from the sensors is obtained, it is sent to the cloud through a base station where the data is processed using various algorithms and a fine-grained air quality pollution map of the city is generated to help understand the air pollution levels properly.*

***Keywords:** LoRa, WAN, Smart City, Sensors, Sensor Network, RTD*

I. INTRODUCTION

Air quality in urban communities is the aftereffect of an intricate connection between normal and anthropogenic natural conditions. Air contamination in urban areas is a genuine ecological issue – particularly in developing nations. The immediate surrounding environment consists of various air pollutants which are present as ambient air quality pollutants in the atmosphere. These pollutants are created because of many reasons. The pollutants arise mainly because of the activities of man. Man has been a major contributor to the creation and presence of these emissions in the atmosphere. These emissions are now threating the very way of life on this planet. The increase in greenhouse gases like CO_2 are a primary reason for the increase in the earth's average temperature. It is also responsible for opening up more and more holes in the earth's ozone layer which is letting in more and more ultraviolet or UV rays into the planet's atmosphere. These UV rays are responsible for diseases like skin cancer and the earth's ozone layer used to prevent such rays from penetrating the earth's atmosphere. These 'gaps' or holes in the ozone layer are now a very serious problem which is a threat to all life on the planet. If the earth's ozone layer is completely destroyed, then the planet is susceptible to powerful cosmic rays which travel across space. Therefore, Air pollutants are a major factor in the deuteriation of human and animal life on the planet. These pollutants are not only present in urban areas like cities but also in towns and villages since industrialization has moved to more developing areas also. This is due to the fact that more and more products are being made and the demand for such products has increased drastically. The main cause of air pollution though are the exhaust fumes coming out of internal combustion engine vehicles which form the majority of all vehicles on the road nowadays. The exhaust fumes coming out of these vehicles contain poisonous gases NO, NO_2, O_3 and O_x. These gases threaten the very livelihood of the people driving them and that of those breathing the fumes.

The main cause of the continuation of the existence of these air pollutants is that there is no means of proper control over the sensing if the air pollutants in the atmosphere. The sensing methods in use now are very low resolution in quality and the sensed data obtained from these nodes are not useful for dynamic analysis of the air quality pollutants. Therefore, a need for a better method of sensing the data is required. The current method of sensor nodes also does not incorporate a fine-grained layer of data mapping to further help analyze and understand the importance of the sensed data. To get a better data map and more accurate sensor data, a more dynamic method of sensor distribution is used. The sensors are placed in crowdsourced automobiles which are employed to provide a more dynamic dataset of the air quality while they move throughout the city. They are housed in a small package to accommodate the most minimum space in the automobile. The sensors are also placed along with a LoRa (Long Range) module. These LoRa modules are used since they can transmit over large distances without any observable loss in data transmitted. The data from the automobile is transmitted to a base station which contains a receiver to receive the data and also an uplink to the server. This method improves the quality of the data obtained and also the data obtained through this method is more finetuned since overlapping data of the same area can be obtained and so a more consistent result can be obtained.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
S. Shriram, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
P. Sai Prasanth, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
Meghna Anand, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

The data from the sensors is sent to a base station and this data is further sent to a server to be gathered and ordered. By using this method, a higher resolution of air quality sensor data is obtained. This data can be used to create a more fine-grained data map of the city.

II. RELATED WORKS

A research paper by Andrii Shelestov discuss the goal of ERA-PLANET to improve the condition of the EUA(European Research Area) which is present in the domain of the earth Observation in accordance with European participation to gather on the Earth Observation (GEO) and the Copernicus constellation. The project had been designed with the plan to develop and improve air quality monitoring systems for urban cities by using the data obtained from the remote sensors and a network of air quality sensors that are used for real time AQ monitoring with high spatial and temporal resolution in a city like Kyiv. Similarly, another paper by Edgar Gabriel discussed pollutant levels with various environments to forecast daily pollutant levels for all platforms of information broadcasting. This paper presents a detailed comparison and understanding of the Hadoop MapReduce and Spark programing models for air quality simulations, guiding future code development for the research groups interested in these analyses. Two use cases have been used, namely (i) to calculate the eight-hour rolling average of pollutants in a restricted defined region; (ii) identifying clusters of sensors showing similar patterns in pollutant concentration over multiple years in the state of Texas.

Another paper by Yue Shan Chang, discuss the application of Big data in air quality analysis and prediction to provide a visual result for estimation and prediction. The number of AQ monitoring devices that are being deployed all over the country are increasing over the years. To estimate and accurately predict AQ, such as PM (particulate matter) 2.5, become an important issue for government to improve people's quality of life. So, here we will propose a semantic ETL (Extract-Transform- Load) framework on cloud platform for AQ prediction. We utilize restful web service as the front-end API to retrieve analyzed data, and finally we exploit browser to show the visualized result to demonstrate the estimation and prediction.

Another research team - Vipul Raheja penned a paper making a case for the use of MiSTIC algorithm to address these issues, compare the use of traditional association rule mining with Salmonellosis disease management, and share new insights on the discoveries. The aim of epidemiology is to identify the causes of disease and comparing them with similar and regionally distinct disease patterns in accordance with the risks to one's heath. The main issue in mining is that a large number of rules are discovered and not all are interesting (due to their inability to conclusively mine spatio-temporal prevalence and disease-causing factors), and some rules may be ignored. A detailed case study elaborated here strongly suggest that while comparing traditional associative rule mining with simple spatio-temporal rule mining that is taken in consideration with spatio-temporal interdependencies in disease data, can provide new, accurate and valuable scientific insights in efficient disease surveillance, prediction and management.

Finally, A paper on Research of time series air quality data based on exploratory data analysis and representation in 2011. Many urban areas monitor air quality in real time to study the effects of the air quality change on the environment and also collect this environmental data in addition to the air quality data. The paper explores the topic and provides insight about the time distribution of air environmental quality and its dynamic changes in real time.

III. PROPOSED SYSTEM

The system consists of sensors, a communication module and a data prediction platform. The sensors are housed in a small package so as to mount it to a crowd source vehicle. This is done so as to maximize the amount of variable data obtained and also by using this method, we can also obtain sensor data which is more localized. This helps in gathering high resolution data which can be used for more creating a more effective data map of the city. This data map can be then used to help understand the pollution hotspots in the city and devise a method to control them. A variety of sensors are used to gather more accurate data from the environment. The different sensors in use are MQ-135(Gas Sensor), DHT11(Humidity and Temperature sensor) and BMP180(Barometric Pressure and Altitude Sensor). These 3 sensors together give a combined reading which is used to help gather a more fine-grained dataset. The sensors are mounted onto a crowdsourced automobile as a small package along with a LoRa module. LoRa is being used in this implementation because of the enormous range that LoRa provides us over current transmission methods. The LoRa modules also go to deep sleep mode when they are not in use, which reduces the power use even more than regular. The regular power draw of the LoRa module when it is use is around 2mA. When it is not in use, it automatically goes into deep sleep mode where it uses around 10uA which is around 0.01mA. So, the entire setup of the sensors and the LoRa transmission modules consumes very minimal power and so causes little burden on the owner of the automobile. This remote node setup can be powered using a small battery and the power draw can be reduced even more by using timers on the sensors and periodic delays in transmission times of the LoRa modules.

After the data is sensed, it is then sent over to a base station by means of a LoRa transmitter. The base station is connected to the cloud by means of an ESP8266 module running micro – python. Since the LoRa transmitters can transmit over large distances, the base station can be kept relatively in the center of the center and all the remote sensors will be able to transmit to the base station at any point of time. The theoretical range of the LoRa transmitters is around 10km which gives us plenty of versatility when deciding the location of the base station. The LoRa used in this system is the LoRaWAN SX1278. The reason behind the use of LoRa in this system is the fact that LoRa is very resistant to both in-band and out-of-band inference mechanisms.

It also operates in an unlicensed band and supports indoor applications like Wi-Fi and it is also highly secure like cellular. LoRaWAN combines the best of both Wi-Fi and cellular to offer an excellent, flexible and economical solution for the problem in the present system.

Once the data is sent to the cloud, the data is collected and a clustering algorithm is implemented onto the data. We use the clustering algorithm in order to group the data together and analyze it. Once the data has been obtained from the sensors through the cloud via the base station, the data exhibits high temporal coherency. A lot of energy is often used when transmitting the data. But since, we are using LoRa based transmitting modules, the energy draw is very minimal. So, there is very less temporal coherency. In order to make some sense of the data collected in the cloud, we have to implement a clustering algorithm onto the collected data. The clustering algorithm that we use in this system is the K-means algorithm. The k-means algorithm is an unsupervised clustering algorithm that uses distance-based measurements to define the similarity of the data points. The K-means algorithm is an iterative algorithm. So, it keeps performing the same loop until a mean deviation of one is obtained. We use the algorithm by first defining the variable k which stands for the number of clusters that is to be assumed. The clusters will be chosen at random and once the algorithm has finished acting on the data points, a data map is plotted which gives us a fine-grained air pollution map of the city

IV. SYSTEM MODULES

Sensing and Interpretation:

The sensing and interpretation module in the system primarily consists of the sensors. The system is equipped with a number of sensors each different from one other and each having their own task to perform. The system is equipped with 3 primary sensors. These primary sensors each detect multiple parameters. The sensors in use in this system are MQ-135, DHT11 and BMP180. The MQ-135 is the gas sensor which detects a number of gases like CO2, CO, Alcohol, Tolueno, Acetona and RO.

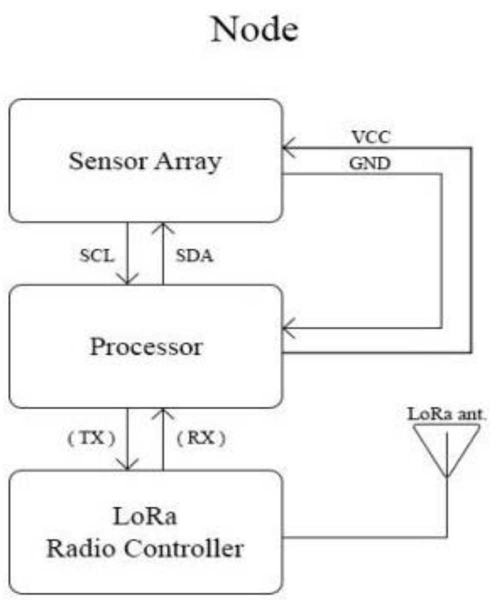

Fig 1.Node Module

The next sensor in use is the DHT11 sensor which detects humidity and temperature. The last and final sensor is the BMP180 which detects the atmospheric pressure in atm and altitude in meters. The sensors are packed in a small package along with a LoRa transmission module. The sensor package in described in fig.1. The LoRa module in use in this system is the SX1278 based RA 01 LoRa module. The LoRa module transmits the data from the sensors on the crowdsourced automobiles to a base station fixed at a central location.

Radio and Communication:

All the communication between the sensors and the base station is done by using the LoRa module. In this system, we use the SX1278 based RA 01 LoRa transmitting module. We use LoRa in this system because unlike other conventional transmitting methods, LoRa doesn't suffer from data losses at large distances. This is because the theoretical range of a LoRa transmission signal is around 10km. Another reason for using LoRa is the availability of a deep sleep mode in the LoRa module.

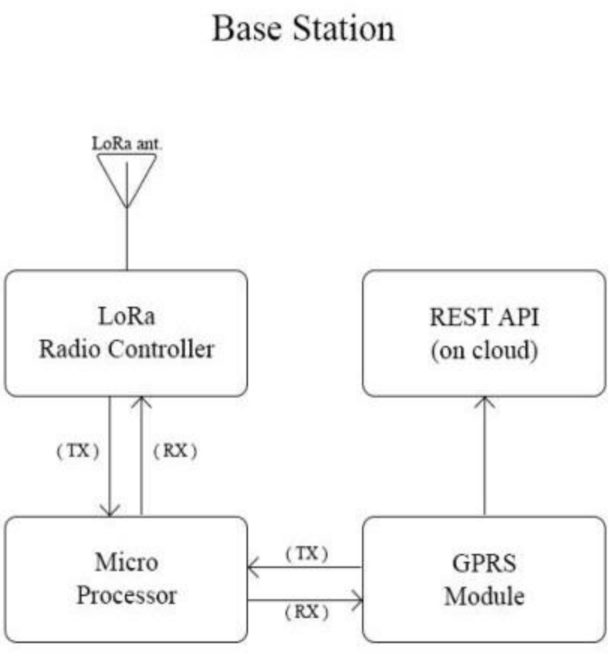

Fig 2.Base Station

The deep sleep mode on the LoRa is activated when the module is not transmitting any data. We can also program the module to go into the deep sleep mode at specific intervels of time between transmissions. The module uses around 2mA when in use and when it is in it's deep sleep mode, it uses around 10uA. The LoRa module sends the data to the base station where an ESP8266 board is connected to the cloud by means of Wi-Fi or a network cable. The system is illustrated in fig.2.

Data Clustering and Prediction:

Once the data is sent to the cloud, we gather up all the data and perform clustering algorithms on the sensed data. By performing a clustering algorithm on the data, we are effectively gathering up all the data and classifying which of the data are useful and which of them are not. The cloud module is illustrated in fig.3.

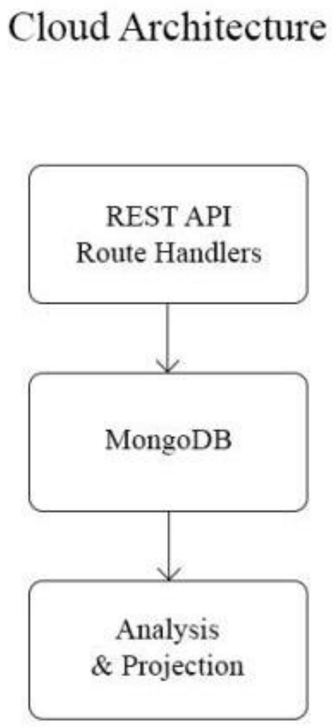

Fig 3.Cloud Architecture

The K-means algorithm can be explained mathematically by means of the following formula:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} (||x_i - v_j||)^2 = 1$$

Where,

$||x_i - y_j||$ is the Eucledian distance between a point, x_i, and a centroid y_j, iterated all k points in the i^{th} cluster, for all n clusters.

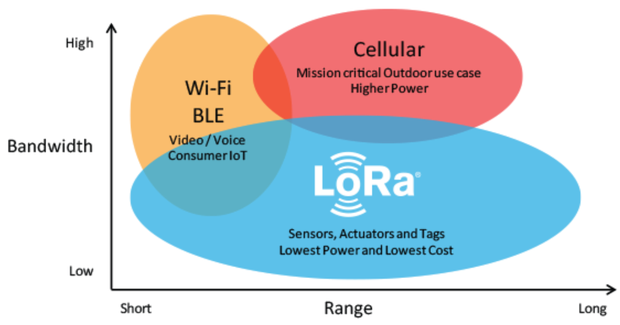

Fig 4. Wireless Network Comparion Graph

V. RESULT

This paper overall solves the problem of low-resolution air quality data by an offline node sensor system that decreases power consumption and cost. It also improves the prediction accuracy by iterative clustering under the k-means algorithm. It's computational efficiency is high along with improving the validity and practicability of the prediction as a whole. The resolution of the output is more dynamic and feasible allowing the end user to have a better, more clear understanding of the air quality of the environment.

VI. CONCLUSION

This system consists of three main components that co-operate in a such a way to provide a dynamic solution to the problem of low-resolution air quality information. The seamless connection between the nodes, base stations and the cloud platform comes together to accurately predict the air quality of respective cities/areas in an accurate way. As soon as the sensor nodes collect air quality data from the environment, they're in, the processor pushes it to our communication medium, LoRa. LoRa provides for a higher range, low bandwidth communication interface between devices in a more effective way. After demodulation and decryption of the data, a microprocessor transmits said data to a GPRS module that uses 2G or 3G to store it on a cloud using REST API. So as to maintain this data for easier organization and understanding, route handlers come into play for optimum functionality. The air quality data in stored in a NoSQL database through MongoDB as it supports the required REST APIs and more. This data is then clustered and by the use of K means algorithm, a prediction is done and then projected onto a heat map. This solves the problem statement by providing the best, most dynamic and efficient way of portraying air quality information to a community of people for their health safety.

FUTURE WORK

Thus, it is evident that there is a need more a better air pollution monitoring system now more than ever. There are several challenges that need to be addressed in this system. The first issue that needs to be improved is the range of the system. We can maximize the range by effectively using satellites to bounce the signal and project the signal as a conical projection rather than a radial projection. By doing so, we can increase the effective transmission range of the LoRa modules exponentially. The next issue that needs to improved on is the over all size of the node, the sensor node package to be exact. The node package can be reduced in size even further by packing the sensor circuit even closer and packing it in a small enclosure. Another expansion that can be implemented on this system in the future is the addition of more and more advanced sensors that can detect minute change in atmospheric gaseous levels. The final improvement to this system is to reduce the overall cost of making the nodes and the base station in order to make it more accessible to all.

REFERENCES

1. W. Yu, "Spatial co-location pattern mining for location-based services in road networks," Expert Systems with Applications, vol. 46, pp. 324–335, 2016.
2. C. Xue,W. Song, L. Qin, Q. Dong, and X.Wen, "A spatiotemporal mining framework for abnormal association patterns in marine environments with a time series of remote sensing images," International Journal of Applied Earth Observation and Geoinformation, vol. 38, pp. 105–114, 2015.

3. H. Nguyen, W. Liu, and F. Chen, “Discovering congestion propagation patterns in spatio-temporal traffic data,” IEEE Transactions on Big Data, pp. 1–1, 2016.
4. Aggarwal and D. Toshniwal, “Spatio-temporal frequent itemset mining on web data,” in 2018 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2018, pp. 1160–1165.
5. G. Atluri, A. Karpatne, and V. Kumar, “Spatio-temporal data mining: A survey of problems and methods,” ACM Computing Surveys (CSUR), vol. 51, no. 4, p. 83, 2018.
6. Y. Shao, B. Liu, S. Wang, and G. Li, “A novel software defect prediction based on atomic class-association rule mining,” Expert Systems with Applications, vol. 114, pp. 237–254, 2018.
7. S. A. Aljawarneh, R. Vangipuram, V. K. Puligadda, and J. Vinjamuri, “Gspamine: An approach to discover temporal association patterns and trends in internet of things,” Future Generation Computer Systems, vol. 74, pp. 430–443, 2017.
8. C.-H. Chee, J. Jaafar, I. A. Aziz, M. H. Hasan, and W. Yeoh, “Algorithms for frequent itemset mining: a literature review,” Artificial Intelligence Review, pp. 1–19, 2018.
9. M. Antonelli, P. Ducange, F. Marcelloni, and A. Segatori, “A novel associative classification model based on a fuzzy frequent pattern mining algorithm,” Expert Systems with Applications, vol. 42, no. 4, pp. 2086– 2097, 2015.
10. N. Aryabarzan, B. Minaei-Bidgoli, and M. Teshnehlab, “negfin: An efficient algorithm for fast mining frequent itemsets,” Expert Systems with Applications, vol. 105, pp. 129–143, 2018.

AUTHORS PROFILE

Shriram S, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India, research interest on AI and Machine Learning.

P. Sai Prasanth, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India, research interest on AI and Machine Learning.

Meghna Anand, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India, research interest on AI and Machine Learning.