Manuscript (Discussion and Conclusion)

Prithvi Yarlagadda

Discussion:

Tried to implement different models to come up with better predictions of air pollution and different kind of air pollutants in the atmosphere. Air pollution is one of the most dangerous pollutions humans are facing right now. Air pollution may result in various harmful diseases and mortality rate will also fall due to this. There are many factors which results in air pollution, humans are also one of the factors for causing air pollution and important factor. Mostly air pollution is caused due to motor vehicles and many recent electronic devices. There are many ways to calculate the air pollution, sensors can be used to understand the percentage of different parameters which cause the air pollution.

These research also considered different factors varying from simple highways to as complicated as emission of gases with respect to speed. These all factors play major role in air pollution. In our research we consider the location, time and different gases as we have access to this data only and perform different regression models and test which one is best suited and gives more accurate results in predicting the pollution for next hour or next day. This paper will also provide with the areas which are highly polluted. This can be used in various field to make the world better place by implementing some measures reduce the pollution or showing the public what are the dangerous places to live or travel regarding air pollution.

Air pollution will have different parameters to consider. Air pollution can be measured using Air quality Index (AQI) metric. This is measured using nearly 449 observation points in total. If AQI increases it also increase chances of health related issues. Sensors can be used for measuring parameters. For the sake of dataset they used one sensor for each traffic sensors available in the area. These sensors will give information and values like Carbon Monoxide levels, Nitrogen dioxide level, Sulphur Dioxide level, Particulate Matter and ozone index level as specified accordingly in Air Pollution Index from wiki. Sensor measures the values by initially assigning it to a value from 25 to 100 according what it is measuring and how dense it is for example Carbon Monoxide. Next for every 5 minutes previous value will be added by a random number from 1 to 10 if its value is below 20. Same way if the value is above 210, a random number from 1 to 10 is subtracted from the previous value. Else a random value from -5 to +5 is added to last value. It is followed this way because the values would not fall in low and high suddenly and keep the values more realistic and confine them in bounds.

Pollution is relatively proportional to the population of the area. For example less densely populated areas have less transportation so this result in less gas emissions and which results in less pollution. So correlation between pollution and population is also considered in study. This makes huge difference in understanding and finding the patterns. If we did not consider the population and calculated the pollution, we cannot rely on that results and it is not a good model to consider for further calculations or predicting the pollution. One limitation for this approach is that we need to have access to the population dataset and we need to have location data too for that population. We have the pollution dataset and locations based on sensors. But we do not have sensors for calculating the population. Even if we had, it would not be same locations as the sensor data. The population and pollution locations will operate on different areas. So this approach cannot be accurate enough, though we can find the city population and can calculate. But this will be for whole and cannot be relied on it.

These research also considered different factors varying from simple highways to as complicated as emission of gases with respect to speed. These all factors play major role in air pollution. In our research we consider the location, time and different gases as we have access to this data only and perform different regression models and test which one is best suited and gives more accurate results in predicting the pollution for next hour or next day.

This paper will also provide with the areas which are highly polluted. This can be used in various field to make the world better place by implementing some measures reduce the pollution or showing the public what are the dangerous places to live or travel regarding air pollution.

Conclusion:

Pollution is when natural resources like Air, Water, Land and other parts of environments starts to become unsuitable or unsafe to use. Air pollution is one of the most dangerous pollutions humans are facing right now. Air pollution may result in various harmful diseases and mortality rate will also fall due to this. There are many factors which results in air pollution, humans are also one of the factors for causing air pollution and important factor. Mostly air pollution is caused due to motor vehicles and many recent electronic devices. There are many ways to calculate the air pollution, sensors can be used to understand the percentage of different parameters which cause the air pollution.

Using different regression models and predicting the next day's or next hour's pollution in that particular area. We can implement all these with different tools available like tableau, R, SQL. As this is not a huge data we can simple load it into R and run regression models. If it is huge data set and even SQL cannot handle the performance, we can go for big data technologies like Hadoop, MapReduce and Spark. If size of dataset is reasonable enough to run on normal databases, it is better to use in it as it is faster and economical for the performance.

The perspectives of this work are understand how transportation technology can help save the environment, decrease vehicles pollution by technology like hybrid cars and electric vehicles, Install sensors and maintain traffic-related air quality updated and Real-time traffic related air pollution can be maintained.