Final Teleprompter script

Contribution of Competitor's Article

My competitor's article "A new approach to link transport emissions and air quality: An intelligent transport system based on the control of traffic air pollution" approach is to analyse the relationship and to develop a more flexible framework to allow communication between transport emissions and air quality concentrations. This paper describes the development of this framework, suggests methodological tools to mitigate its problems and shows its application to the mega-city The result of implementing this methodology would be a system providing high time/space resolution measurements of both air pollutant concentrations and traffic emissions data, as well as real-time transportation and dispersion modelling of those data. The key advantage of the system proposed would be the runtime integration of modelling, to interpret the data measured, with measurements, to validate the data modelled. The findings from the case-study of Beijing show that the integrated system can link traffic air pollution measurements through various modelling modules in order to automate transport-related air pollution assessment.

In this study they followed some methodologies to predict the pollution in urban air pollution. Also explained about how vehicle transportation is important and what are the minimum requirements used for these kind of transportations. By this how our civilization in advancing and improving technological advancements. On the other hand these advancements are also have negative impacts on the world we live in. These negative impacts will result in the environment changes, health concerned issues and our world, earth we live in will day by day will become unsuitable for humans. These are serious issues every nation and every individual should consider. Our future generation will depend on the actions we take now. So considering seriousness of the environment, this study helps in understand distribution and level of air pollution in small area compared to the world Brasov city. They tried different regression models to come up with better prediction of the air pollution. In my research and I am going to consider these methods to understand how it is distributed and predicted and I will implement those on our dataset. I also consider the parameters like correlation between area and pollution, population and pollution. Because we cannot say that these part of city have high pollution based on parameters, there could be more vehicles running in these areas at some particular time interval and which results in these kind of increase in air pollution. So population should also be a main factor in deciding and predicting the air pollution.

Description of Your Contribution

From the dataset we can find the locations and their respective gas emissions. From these factors we can find some patterns like correlation between time and pollution. We can also perform time-series calculations and can predict the pollution on that particular location at this particular interval of time. From this analysis we can mark the highly polluted areas and may avoid frequent visits in these related areas. These can be calculated by the dataset we have, there are 449 locations and each location have one separate file. Each file has around 18000 records with location and emission values. So each location can be taken at a time and sum all the values. These can be categorized into different time series like calculate for monthly, daily and hourly. From this we can find a pattern and may predict for the next day or hour. These can be very helpful in understanding the pollution distribution and help people in avoiding those areas. We can use Brasov city data and

use the location data and find the values of the pollution and show case the highly polluted areas in the city by heatmap. Using these heatmap we can find the highly polluted areas so that public can avoid those in case of travelling. This can also help in taking measures for reducing the pollution in those areas, may be by planting more trees or making vehicles to take alternate route. Some ideas of show casing the polluted areas are by making specific areas green where it is not polluted and red where it is highly polluted.

So each location can be taken at a time and sum all the values. These can be categorized into different time series like calculate for monthly, daily and hourly. From this we can find a pattern and may predict for the next day or hour. These can be very helpful in understanding the pollution distribution. Used Scipy.signal.coherence library for the cross-spectrum analysis.

Data Source and Content

The dataset is from Citypulse website, where it provides with the air pollution dataset of Brasov in Romania. There are 449 different locations where data is collected. Dataset consists of 449 different csv files, each file for each location. Each csv file have parameters like location (latitude and longitude's), timestamp (it is recorded for every 5 minutes), ozone level, particulate matter level, carbon monoxide level, sulphur dioxide and nitrogen dioxide. Dataset is from the dates August 1 2014 to September 30 2014, 2 months of data which is recorded every 5 minutes. Using this dataset we can find some pattern in the air pollution and help in bringing that down. There are some ideas or implementations which we follow through in this paper for better understanding and conclude with some insight. Road traffic data and chemical pollution data. The used method for this study was the traffic data manual collecting. For one input it is established the following data measurement: traffic volume, vehicles number for each direction (left, forward, right) and for each vehicle category [2]. The traffic flow volume was determined by counting the vehicles total number for one intersection during one rush hour (Example: 8.00-9.00, 15.00-16.00). Chemical pollution measurement methodology: for each studied intersection were selected a number of measurement points for chemical pollution data. The measurement points number was chosen according to the traffic flows, the intersection geometrical parameters and placement of the buildings.

Your Method

I used Apache Zeppelin to run different methods in our project. Apache Zeppelin enables interactive data analytics on a web based notebook. Apache Zeppelin supports different languages like Python, Scala, R, SQL and more. The main advantage of using Apache Zeppelin is, it has built in Apache Spark integration. In many other notebook or tools we need to build separate module or import the libraries required. Some of the features Apache Zeppelin provides with spark integration are Automatic injection of data using SparkContext and SQLContext, load libraries or jar files from local file system or maven repositories, display the job progress and job cancelation. Normally Apache Zeppelin comes with whole package required for data analysis. For example, matplotlib library is mostly used for the data visualization and data analysis in python, so latest version of Apache Zeppelin comes with integrated Matplotlib library, we do not have to import it separately and also leverage the Spark performance speed by pyspark Interpreter which fast and reliable. In the first step of data analysis, linear relationships between analyzed pollutants and meteorological parameters were tested using Pearson's correlation resulting 48 possible pollutant –meteorological factor pairs.

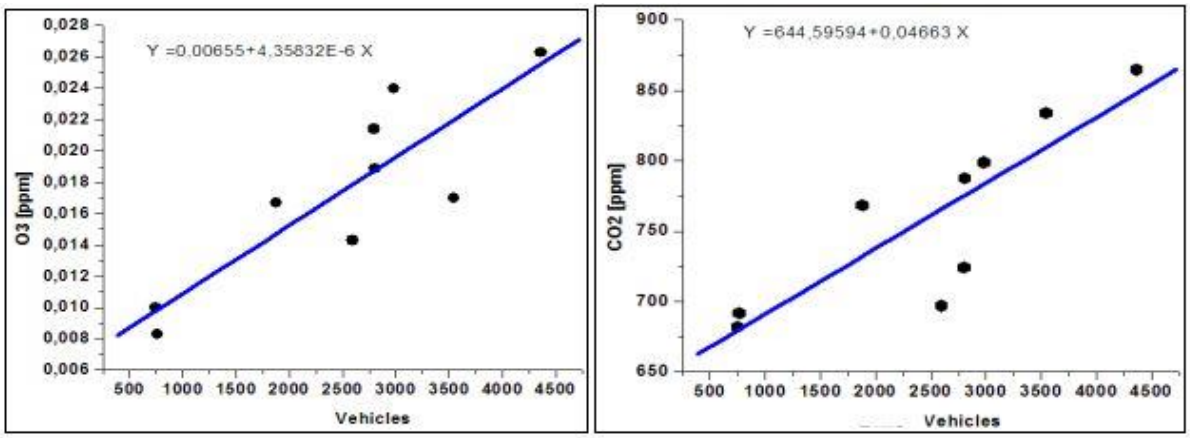
Quantitative Results 1

I tried different prediction models to come up with predicting air pollution in Brasov city. We can use this models to predict pollutions of different cities. These predictions are also based on the

parameters like correlation between area with pollution and population with pollution. Because we cannot say that these part of city have high pollution based on parameters, there could be more vehicles running in these areas at some particular time interval and which results in these kind of increase in air pollution. So population should also be a main factor in deciding and predicting the air pollution.

Parameters	*Global Radiation*	*Relative Humidity*	*Temperature*
BV-2 urban station			
Sulfur dioxide - SO_2 (n = 2721)	-0.008	0.196**	0.008
Nitrogen dioxide - NO_2 (n = 2911)	0.067**	-0.076**	-0.194**
Particulate Matter PM2.5 (n = 1730)	0.077**	-0.055*	0.047*
BV-3 traffic station			
Sulfur dioxide - SO_2 (n = 2832)	-0.104**	0.093**	-0.016
Nitrogen dioxide - NO_2 (n = 2832)	0.041*	0.021	-0.102**
Particulate Matter PM10 (n = 2832)	-0.014	0.098**	-0.170**
BV-5 industrial station			
Sulfur dioxide - SO_2 (n = 2679)	0.001	0.024	-0.003
PH-3 suburban station			
Sulfur dioxide - SO_2 (n = 3066)	0.088**	-0.050**	0.026
Particulate Matter PM10 (n = 3066)	-0.163**	0.207**	-0.453**
PH-4 industrial station			
Sulfur dioxide - SO_2 (n = 3265)	0.336**	-0.206**	-0.001
Nitrogen dioxide - NO_2 (n = 3265)	-0.169**	0.067**	-0.348**
PH-6 industrial station			
Sulfur dioxide - SO_2 (n = 3284)	0.185**	0.207**	-0.202**
Nitrogen dioxide - NO_2 (n = 3284)	-0.112**	-0.081**	-0.054**
DB-1 industrial station			
Sulfur dioxide - SO_2 (n = 957)	0.143**	-0.117**	0.119**
Nitrogen dioxide - NO_2 (n = 1221)	-0.315**	0.087**	-0.204**
Particulate Matter PM10 (n = 957)	-0.036	-0.008	-0.329**

Quantitative Results 2



The most common frequencies were 0.0416 and 0.0833. An interesting remark is that the week period (7 days) did not occur in the highest five peaks of cross amplitude. 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. 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.

Discussion: Comparison With Your Competitor

My competitor's paper "A new approach to link transport emissions and air quality: An intelligent transport system based on the control of traffic air pollution" main approach is to build a regression model based on pollution data, Cross-spectrum analysis of air pollution and weather data, Reveal the correlations between two series at different frequencies, The number of valid observations was odd, padding has been applied to have even series for Fourier analysis. Whereas my approach is more of an analytical approach towards the prediction by building regression model by considering parameters like traffic volume and air pollutants, Correlation between air pollutants and weather parameters, 33 strong correlations ($p < 0.01$) and 3 moderately strong correlations ($p < 0.05$) were identified. The remaining 12 pairs did not show a linear correlation.

An important element of this effort is the development of alternative powertrains to the internal combustion engine (ICE). While a number of these alternatives present a great expectation for improving energy efficiency or reduced emissions, some early technical solutions lack the power density of ICEs and deficiency implies that either performance must be compromised or the rest of the vehicle must be made lighter. Different studies show that the internal combustion engines (ICEs) have the potential to increase the fuel efficiency by 2020. The alternative powertrains can be included into four main categories which can be easily developed: hybrids, fuel cell, battery powered, and the compressed natural gas. When different energy carriers with varying degrees of energy losses during fuel production are compared and distribution is used, primary energy efficiency analysis becomes necessary. The European economy loses nearly 100 billion euros, or 1% of the EU's GDP per year. Year by year the level of air and noise pollution is increasing. Urban traffic is responsible for 40% of CO_2 emissions and 70% of emissions of other pollutants arising from road transport. A solution could be the smaller, efficient and clean vehicles used for local distribution. Negative impacts of long distance freight transport passing through urban areas should be reduced through planning and technical measures.

Performance on Big Data: Time Measurements

The total time taken for execution of my capstone project is around 4 minutes. Digging deep into each phase of my project, data load which I used for this project consists of 3 different data files as I mentioned it in my data and content section. So each file have nearly 500 records and around 10 different columns. These data sets are loaded into python by using csv funtions. Each file loading took around 20 seconds, so there are 3 data files it took around 60 seconds for data load. Now that we loaded the data, my capstone project is based on these 3 different data sets merging together and running models and finding any correlations between them, so I have to merge these 3 datasets into one dataset and run my models and correlation functions. Data merging is done based on some common columns in both the data. It's just like joins in SQL. So this merging based on some common ground took me around 30 seconds. I have used different predictive models, time series analysis and correlations between different columns not only from same data set but also from different data set like weather and traffic dataset. These two are not related in whole point of view but there should be a small correlation. So these kind of analysis we help in data analytics and make a huge difference in predicting future traffic or weather in that matter. So these analysis took around 2 minutes. This looks like small amount of time, but most of the time consuming part will be analysing data and understanding what to do and how to do it.

Conclusion

The results show an important contribution of short-term fluctuations to the total variance for analyzed pollutants (nitrogen dioxide, sulfur dioxide and suspended particles), with high dependence on the meteorological factor (incident radiation, air temperature or relative humidity) and less on the site specific conditions. This type of exploratory approach might help as a pre-assessment tool identifying constants or patterns that can characterize the interactions between variables, which is a useful tool for the parameterization and calibration of air pollution models. Furthermore, the technique can provide comprehensive classifications of the monitoring sites.