Memo about $PM_{2.5}$ in Beijing: For Those Who May Concern

Xiaozhu Zhang

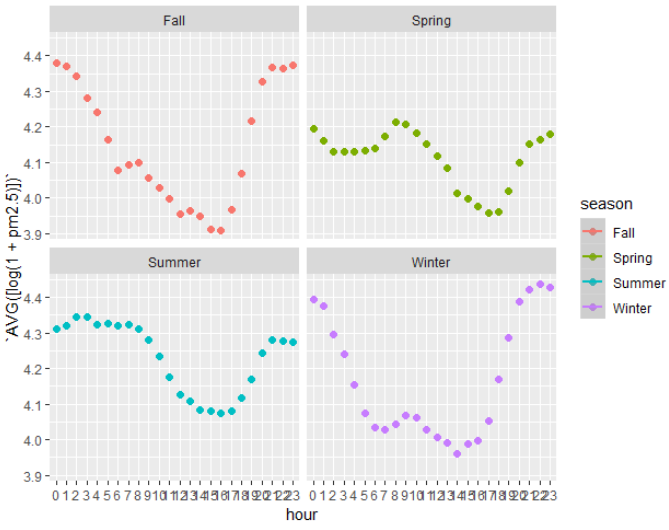
(This is a memo for management-level audience who may concern about the air pollution in Beijing, China. You do not need to be an expert in data science to understand this memo, so feel free to read it! However, if you want to learn more technical details of this project, please read the detailed report from which this memo is abstracted.)

Before the Beijing 2008 Summer Olympic Games, the severe air pollution in Beijing had already caught the whole world's attention. Among all those gauges of air pollution, $PM_{2.5}$ concentration is no doubt the most famous and effective. According to U.S. Environmental Protection Agency (EPA), the air quality index (AQI) could be regarded as "Good" if AQI is less than 50, whereas the average AQI in Beijing from 2010 to 2014 was almost 100. In order to deal with this serious problem, we use dataset from UCI Repository (with time series data from 2010 to 2014) to explore the factors influencing $PM_{2.5}$ in Beijing, perform machine learning algorithm to impute missing values, and finally comment on the trend of $PM_{2.5}$ during the 5 years studied.

Relationship between seasons, hours and $PM_{2.5}$

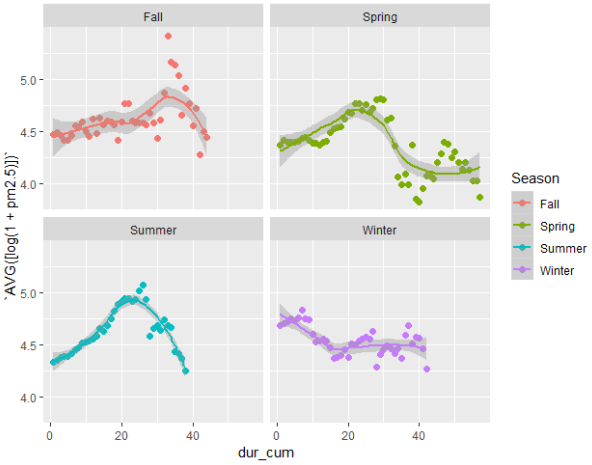
Beijing is located in the temperate region of mason climate, enjoying 4 fairly divergent seasons. Both temperature and humidity could influence the concentration of $PM_{2.5}$ significantly; however, human activity is something we should never ignore. Adjacent to Shanxi Province, which is the biggest and most famous coal-producing region in China, citizens in Beijing always burn a huge amount of coal for heating every winter and therefore pollution could increase dramatically.

According to the figure on the right, regardless of the seasons, $PM_{2.5}$ concentration will fall during 1am to 6am, rise during morning rush hours, fall again in the afternoon, and rise again after 6pm. This specific pattern could be explained by temperature gap between day and night, and also human activity such as traffic jam and residential heating. We notice that in Spring and Summer the peak around morning rush hours is almost at the same level as that after 6pm, whereas in Fall and Winter the peak after 6pm is much higher. This observation gives a hint that residential heating is more polluted than residential cooling in Beijing, and also more polluted than industrial cooling and heating.



Relationship between wind and $PM_{2.5}$

Take a quick look at the topography of Beijing, we will easily find that there are Yanshan Mountain and Taihang Mountain in the north and west of Beijing, blocking the entry of water vapor; while in the south and east of Beijing, there is a large plain connecting Beijing to the coastal area. In fact, wind from southeast is more common in Summer while wind from northwest is more common in Winter, indicating that Summer in Beijing is wetter while Winter is drier.



Comparing figure above (NW wind) and figure on the left (SE wind), we could observe a relatively opposite pattern. With northwest wind persisting, $PM_{2.5}$ concentration will first fall and then arise; while with southeast wind persisting, $PM_{2.5}$ concentration will first arise and then fall. It is not easy to come up with a reasonable explanation for this phenomenon, but here is a guess: dry wind is more likely to blow away tiny particles such as $PM_{2.5}$, while wet wind is more likely to bring about fog or smog; however, many factories in heavily polluted industries are located in the north and west of Beijing, while areas in the south and east of Beijing are more clean. These two opposite factors intertwine with each other, and the wet / dry factor works before the polluted particle factor; in other words, if the wind duration is long enough, opposite phenomenon would appear.

Relationship between precipitation and $PM_{2.5}$

Linear regression model is built to analyze the relationship between precipitation and $PM_{2.5}$. The concentration of $PM_{2.5}$ will increase if there is precipitation, and with precipitation increasing, the increment will arise first and then fall, reaching almost zero when the precipitation is the maximal. Although it is anomalous to get positive parameters, we do conclude that a heavy rain/snow will refresh the air, compared with light rain/snow.

Relationship between other meteorological factors and $PM_{2.5}$

In fact, Dew Point, Temperature and Pressure are highly correlated. Surprisingly, although the correlation coefficients between $PM_{2.5}$ and other variables are quite small, they are significant in the sense of p-value (a term in Statistics). It will occur sometimes especially when the total sample size is pretty large.

In general, when the dew point and temperature increase while pressure and wind speed fall, $PM_{2.5}$ concentration is more likely to increase.

How to predict $PM_{2.5}$?

The Random Forest algorithm is definitely a good way to predict $PM_{2.5}$ concentration. If you input the value of Dew Point, Temperature, Pressure, logarithm of Cumulative Wind Speed (plus 1), the specific time and season of the observation, wind direction and precipitation level into the model we have already built, you

will certainly get the logarithm of $PM_{2.5}$ (plus 1). In fact, our model enjoys an accuracy of 64.41%, which is pretty high when it comes to predicting a value in the natural world. The machine learning model we built could not only make predictions, but also conduct imputation of missing values in the original dataset.

Was air quality in Beijing getting better from 2010 to 2014?

Although the trend is extremely slight, the answer is yes. By the non-parametric method "Cox-Stuart test", we concluded that $PM_{2.5}$ concentration is slowly decreasing in the 5 years studied in our project.

What can we do to improve air quality in Beijing?

Inspired by the abundant conclusions we have reached, there are a lot of actions we could take.

1. Encourage citizens in Beijing to travel more by public transportation instead of private cars, especially during rush hours.
2. Provide subsidies for clean energy for residential heating and cooling, such as natural gas; at the same time, reduce the mining and consumption of coal, especially in Shanxi Province.
3. Adjust taxes on factories based on the energy used: the cleaner the energy is, the lower the taxes will be.
4. Move factories in north and west of Beijing to the place with downwind direction.

With actions and efforts, we believe that air quality in Beijing must get even better in the future!