The Study of Relationship Between Air Pollution and Negative Sentiment with Google Trend

Kim Minseob   
*School of Computer Science and Engineering  
Kyungpook National University*Daegu, Korea  
dmg04158@gmail.com

*Abstract*—The correlation with weekly average CAI (air pollution index) and Google Trend data with given negative keywords are calculated and evaluated to show positive relationship between air pollution and negative sentiment in Korea.

Keywords—correlation, air pollution, sentiment analysis, Google Trend

# Introduction

The harms on human health by air pollution is known but detailed sentiment analysis is not conducted in the past.

In this study, the air pollution is calculated into CAI (Comprehensive Air-quality Index) and compared with Google Trend data with given negative keywords that reflects the Korean people’s interest in body and mental health.

The data is evaluated with correlation values and shows slightly weak relationships.

# Air Pollution Data Representation

## Data Source

Air Korea, a public website by KECO (Korean Environment Corporation) announces air pollutants concentrations and air quality score hourly. The measured air pollutants are: particulate matter in PM10, PM2.5, Ozone, Nitrogen Dioxide, Carbon Monoxide and Sulfur Dioxide.

The available daily average measurement of 6 pollutants were January 2014 to October 2021. In this study, the selected data were January 2021 to October 2021 and measured in Jung-gu, Seoul. The missing datas are interpolated.

## Comprehensive Air-quality Index

The air pollution is related with all 6 pollutants above. To determine the air quality in numerically, CAI (Comprehensive Air-quality Index) is needed. CAI is only used in Korea and is an alternative form of AQI (Air Quality Index) which is used in USA. It represents air quality in numerical value and equivalent categories. Table I shows how CAI is represented.

1. CAI Representations

| Categories | Index value range |
| --- | --- |
| Good (좋음) | 0-50 |
| Moderate (보통) | 51-100 |
| Unhealthy (나쁨) | 101-250 |
| Hazardous (매우나쁨) | 251-500 |

To yield index value, the CAI equation is applied for all 6 pollutants regarding their ppm or measurement.

is the measured value of one of the pollutants. The , , and values can be looked-up in the table with by which category it belongs to.

1. Breakpoint value table

| ***Pollutant type*** | ***Category*** | value range |  |  |
| --- | --- | --- | --- | --- |
| PM10  () | Good | 0-30 | 0 | 30 |
| Moderate | 31-80 | 31 | 80 |
| Unhealthy | 81-150 | 81 | 150 |
| Hazardous | 151-600 | 151 | 600 |
| PM2.5  () | Good | 0-15 | 0 | 15 |
| Moderate | 16-35 | 16 | 35 |
| Unhealthy | 36-75 | 36 | 75 |
| Hazardous | 76-500 | 76 | 500 |
| Ozone (ppm) | Good | 0-30 | 0 | 30 |
| Moderate | 31-80 | 31 | 80 |
| Unhealthy | 81-150 | 81 | 150 |
| Hazardous | 151-600 | 151 | 600 |
| Nitrogen Dioxide (ppm) | Good | 0-0.03 | 0 | 0.03 |
| Moderate | 0.031-0.06 | 0.031 | 0.06 |
| Unhealthy | 0.061-0.2 | 0.061 | 0.2 |
| Hazardous | 0.201-2 | 0.201 | 2 |
| Carbon Monoxide (ppm) | Good | 0-2 | 0 | 2 |
| Moderate | 2.1-9 | 2.1 | 9 |
| Unhealthy | 9.1-15 | 9.1 | 15 |
| Hazardous | 15.1-50 | 15.1 | 50 |
| Sulfur Dioxide (ppm) | Good | 0-0.02 | 0 | 0.02 |
| Moderate | 0.021-0.05 | 0.021 | 0.05 |
| Unhealthy | 0.051-0.15 | 0.051 | 0.15 |
| Hazardous | 0.151-1 | 0.151 | 1 |

When the category is determined during the look-up of and , the and values can be looked-up in the same manner.

1. I value table

| ***Pollutant type*** | ***Category*** |  |  |
| --- | --- | --- | --- |
| Any | Good | 0 | 50 |
| Moderate | 51 | 100 |
| Unhealthy | 101 | 250 |
| Hazardous | 251 | 500 |

After preparing all the values needed, the  value of the equation can be yielded. To get the final CAI value which represents the current comprehensive air quality, those conditions should be followed.

* If there is 0 or 1 pollutant that categorized as ‘Unhealthy or ‘Hazardous, the final CAI value is .
* If there is 2 pollutants that categorized as ‘Unhealthy or ‘Hazardous’, the final CAI value is .
* If there is 3 or more pollutants that categorized as ‘Unhealthy’ or ‘Hazardous’, the final CAI value is .

## CAI Preparation

As Google Trend shows only weekly search trend, the calculated CAI value were converted into weekly average CAI value. The calculation is executed with Python.

1. Weekly Average CAI

# Google Trend Data Representation

Google Trend shows how much a keyword was searched on Google during specific period. Each data point holds linear score of 0 to 100 that combined weekly, especially Sunday to Saturday. The maximum searched count is transformed as 100. The other searched count are transformed linearly proportional to the maximum searched count. Same as CAI, the selected period is between January 2021 to October 2021.

The keywords used in this study are: ‘불면증’ (insomnia), ‘어지러움’ (dizzy), ‘만성피로’ (chronic fatigue). ‘우울증’ (depression), ‘우울해’ (depressed), ‘자살’ (suicide), ‘열등감’ (sense of inferiority), ‘외로움’ (loneliness).

1. Google Trend Keyword ‘불면증’
2. Google Trend Keyword ‘어지러움’
3. Google Trend Keyword ‘만성피로’
4. Google Trend Keyword ‘우울증’
5. Google Trend Keyword ‘우울해’
6. Google Trend Keyword ‘자살’
7. Google Trend Keyword ‘열등감’
8. Google Trend Keyword ‘외로움’

# Combined Data

The weekly average CAI and Google Trend are combined and represented as correlation values. Considering that the CAI value is inconstant during the first half of the year, while constant during the latter half, the correlation value is represented again regarding only the first half. The results are shown in Table IV.

1. Correlation Between CAI and keywords

| Keyword | Correlation  (overall) | Correlation (first half) |
| --- | --- | --- |
| 불면증 (insomnia) | 0.1864 | 0.0418 |
| 어지러움 (dizzy) | -0.2262 | 0.2837 |
| 만성피로 (chronic fatigue) | -0.0750 | 0.3558 |
| 우울증 (depression) | 0.1659 | 0.0011 |
| 우울해 (depressed) | 0.0047 | 0.2798 |
| 자살 (suicide) | 0.1739 | -0.0471 |
| 열등감 (sense of inferiority) | 0.2081 | 0.2232 |
| 외로움 (loneliness) | -0.0064 | 0.1966 |

* The keyword ‘불면증’ shows weak association overall.
* The keyword ‘어지러움’ shows weak association in the first half.
* The keyword ‘만성피로’ shows slightly strong association in the first half.
* The keyword ‘우울증’ shows weak association overall.
* The keyword ‘우울해’ shows weak association in the first half.
* The keyword ‘자살’ shows weak association overall.
* The keyword ‘열등감’ shows weak association overall and in the first half.
* The keyword ‘외로움’ shows weak association in the first half.

# Conclusion

The used air pollution data is limited in Jung-gu, Seoul, taking that more than 50% of total population of Korea is in Seoul, and the air pollutants have the tendency to move around nearby cities.

Also Google Trend is measured across all territories of the Korea, the population is considered as a strong variable and this meets the condition of the selection of air pollution data.

The results in correlation values are showing rather weak associations but consistent positive relationship either in overall or the first half of the period.

##### References

1. Air Korea, <https://www.airkorea.or.kr/>, final measurement data, the Comprehensive Air-quality Index equation.
2. e-나라지표, [https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx\_cd=1007#](https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1007), the ratio of Seoul population.