

AIRSCAPE (project idea)

Oleg Bondarenko, 22.09.2021, <https://protw.github.io/oleghbond>

Problem

The problem with instrumental methods of air quality monitoring is that

1. it is virtually impossible to provide the required spatial density of sensors;
2. the sensors can measure a rather limited list of physical factors that determine the air quality, so, the resulted AQI is very often incomplete and, therefore, inadequate;
3. overcoming the problems in pp. 1 and 2 requires the deployment of additional infrastructure and its maintenance and therefore significant resources.

Solution

We would like to use the Social Media Data (SMD) for assessment of air quality, particularly, by building a so-called *airscape*.

Therefore, the approach that we propose is an extension of traditional instrumental methods of air quality monitoring and has certain advantages over the latter.

The main advantage of the proposed approach are:

1. a much wider and denser coverage of the geographical distribution,
2. a more comprehensive list of possibly assessed physical factors
3. no need to deploy additional instrumental infrastructure on the ground.

Thus, we consider our project as a feasibility study of an extension of the air quality monitoring system.

Plan

We plan to build the following technology chain:

1. Collecting the SMD (posts) by certain criteria, first of all, geotag (certain area, city), time period, key words presence, other.
2. Analysis of the collected SMD (possibly using Natural Language Processing) in order to get forming factors and resulting emotions.
3. The results of SMD analysis will be mapped with corresponding instrumental measurements of air quality data by means of neural network learning in order to utilise the SMD analysis independently where and when results of instrumental methods are not available.
4. Implementation of results for various scenarios of practical use.

We are going to use posts and, possibly, Likes. By analysing the text of posts we can determine air quality factor (PM, various smells, humidity, transparency, other) and human perception of this factor - at least: positive, neutral, negative.

As a result, we plan to obtain air quality factors connecting to geo location and period of time. So the results can be presented as dynamic maps or something similar.

Results

The result of the project can be used by citizens of big cities and towns, local governments of, first of all, big cities or heavy industrial regions as an indicative early warning system to draw higher attention and undertaking timely countermeasures. The proposed system is an extension of traditional instrumental methods of air quality monitoring.

Based on determined physical / chemical factors that shape air quality, as well as geolocation and the posting time, we plan to present the results in the form of a map - *airscape*. Among the aggregated data, it makes sense to represent the number of publications selected in a particular spatial area, on the basis of which the assessment was made. This is necessary to understand the representativeness of the result.