Ergonomic system of forecasting and analysis of the probability of atmospheric air pollution by emissions of enterprises in the conditions of technogenic development of the world

Alexander Kuzmenko*, Rodion Filippov, Lyudmila Filippova, Maxim Terekhov, and Pavel Vershkov

Bryansk State Technical University, Bul'var 50-letiya Oktyabrya 7, Bryansk, 241013, Russia

Abstract. The article is devoted to the problem of air pollution and its impact on human health and the environment. The paper considers methods of air pollution analysis based on the use of neural networks, taking into account the variety of data from the Internet. The authors emphasize the different effects of pollutants depending on the type, duration and level of exposure, as well as other factors, including individual risks to human health and the combined effects of different pollutants and stress factors. Special attention is paid to the two most common types of air pollution - smog and soot. The uneven distribution of the negative effects of air pollution, which are most often felt in low-income and colored communities, as well as the Air quality Index (AQI) developed by the Environmental Protection Agency, which informs the population about the current state of atmospheric air and its impact on human health, are considered separately. As a result of the work, the criteria for analyzing air quality, including pollution parameters and weather conditions, are presented, and the structure of future data is developed.

1 Introduction

The effects of air pollution on the human body vary depending on the type of pollutant, duration and level of exposure, as well as on other factors, including individual risks to human health and the cumulative effects of multiple pollutants or stress factors.

These are the two most common types of air pollution. Smog (sometimes called ground ozone) appears when emissions from burning fossil fuels react with sunlight. Soot (also known as solid particles) consists of tiny particles of chemicals, soil, smoke, dust, or allergens - in the form of gas or solid particles - that are carried in the air. The sources of smog and soot are similar [1,2,3].

Smog can irritate eyes and throat, as well as damage lungs, especially in children, elderly people and those who work or play sports outdoors. Even worse it is for people suffering from asthma or allergies: these additional pollutants can exacerbate their symptoms and cause...
asthma attacks. The smallest particles of soot in the air, gaseous or solid, are especially
dangerous because they can enter the lungs and bloodstream and worsen bronchitis, cause
heart attacks and even accelerate death. In 2020, a report by the Harvard TH Chan School of
Public Health showed that the death rate from COVID-19 in areas with more soot pollution
was higher than in areas with even slightly less.

2 Materials and methods

Because highways and polluting facilities have historically been located in or near low-
income areas and communities, the negative effects of this pollution have been
disproportionately felt by people living in these communities. In 2019, the Union of
Concerned Scientists found that the impact of soot on Asian Americans is on average 34
percent higher than on other Americans. For blacks the exposure rate was 24 percent higher;
for Latinos it was 23 percent higher. To inform the public about the current air quality, EPA
(US Environmental Protection Agency) [4,5] developed an Air Quality Index (AQI) [3,6],
which measures five most common air pollutants. This index allows informing the population
about the current state of atmospheric air and its possible impact on human health. After the
analysis of the subject area, the following criteria have been identified, according to which
the analysis of air pollution is carried out.

Criteria:
- SO2 readings in micrograms/m3 per hour.
- NO2 readings in micrograms/m3 per hour.
- O3 readings in micrograms/m3 per hour.
- CO readings in micrograms/m3 per hour.
- PM10 readings in micrograms/m3 per hour.
- PM2.5 readings in micrograms/m3 per hour.
- Humidity reading for the current hour.
- Pressure reading for the current hour.
- Temperature reading for the current hour.
- Wind direction for the current hour.
- Wind speed for the current hour.
- Weather forecast for the next hour (7-11items).
- AQI for the next hour.

The structure of future data is also developed (Table 1). Each of the above criteria has
several possible answers. Each answer option has its own value.

<table>
<thead>
<tr>
<th>Criterium</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 SO2 (microgram/m³)</td>
<td>int</td>
</tr>
<tr>
<td>2 NO2 (microgram/m³)</td>
<td>int</td>
</tr>
<tr>
<td>3 O3 (microgram/m³)</td>
<td>int</td>
</tr>
<tr>
<td>4 CO (microgram/m³)</td>
<td>int</td>
</tr>
<tr>
<td>5 PM10</td>
<td>int</td>
</tr>
<tr>
<td>6 PM2.5</td>
<td>int</td>
</tr>
<tr>
<td>7 Humidity</td>
<td>int</td>
</tr>
<tr>
<td>8 Pressure</td>
<td>int</td>
</tr>
<tr>
<td>9 Temperature (kelvin)</td>
<td>int</td>
</tr>
<tr>
<td>10 Wind Direction</td>
<td>from 0 to 360</td>
</tr>
<tr>
<td>11 Wind Speed</td>
<td>int</td>
</tr>
<tr>
<td>12 AQI</td>
<td>from 1 to 10</td>
</tr>
</tbody>
</table>
3 Results

When preparing the data set, the most suitable data were searched for in free access. A data set of air pollution indicators in Seoul from 25 stations for 2019 was selected [7,8], the readings in the set were in different units of measurement and for the accuracy of the neural network, all indicators were reduced to a microgram/meter³.

Python tools were used to check the data for incorrect values (Figure 1).

![Fig. 1. Deleting null values in a dataset [4,5].](image)

All indicator values were brought to AQI (Table 2).

**Table 2.** Bringing the values of harmful substances to AQI [4,5].

<table>
<thead>
<tr>
<th>Index</th>
<th>Ozone (microgram/m³)</th>
<th>PM10 particles (microgram/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0–33</td>
<td>0–16</td>
</tr>
<tr>
<td>2</td>
<td>34–66</td>
<td>17–33</td>
</tr>
<tr>
<td>3</td>
<td>67–100</td>
<td>34–50</td>
</tr>
<tr>
<td>4</td>
<td>101–120</td>
<td>51–58</td>
</tr>
<tr>
<td>5</td>
<td>121–140</td>
<td>59–66</td>
</tr>
<tr>
<td>6</td>
<td>141–160</td>
<td>67–75</td>
</tr>
<tr>
<td>7</td>
<td>161–187</td>
<td>76–83</td>
</tr>
<tr>
<td>8</td>
<td>188–213</td>
<td>84–91</td>
</tr>
<tr>
<td>9</td>
<td>214–240</td>
<td>92–100</td>
</tr>
<tr>
<td>10</td>
<td>≥ 241</td>
<td>≥ 101</td>
</tr>
</tbody>
</table>

The presented part of the table shows the dependence of Ozone index (O3) and the particle (PM 10) on AQI. The same matching was done for all pollutants.

In addition to emission indexes, the work takes into account weather conditions affecting the spread of pollutants. Weather conditions are taken from open sources for 2019 in Seoul and presented in the form of a table [5].
Table 3. Example of weather data in Seoul [5].

<table>
<thead>
<tr>
<th>Humidity</th>
<th>Pressure</th>
<th>Temperature</th>
<th>Wind Direction</th>
<th>Wind Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>807</td>
<td>284.63</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>76</td>
<td>849</td>
<td>284.629</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>76</td>
<td>890</td>
<td>284.627</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>77</td>
<td>932</td>
<td>284.625</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>78</td>
<td>973</td>
<td>284.6229</td>
<td>47</td>
<td>2</td>
</tr>
<tr>
<td>78</td>
<td>1015</td>
<td>284.6209</td>
<td>61</td>
<td>2</td>
</tr>
<tr>
<td>79</td>
<td>1027</td>
<td>284.6188</td>
<td>75</td>
<td>2</td>
</tr>
<tr>
<td>79</td>
<td>1015</td>
<td>284.6168</td>
<td>89</td>
<td>1</td>
</tr>
<tr>
<td>80</td>
<td>1014</td>
<td>284.6147</td>
<td>102</td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td>1013</td>
<td>284.6127</td>
<td>116</td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td>1012</td>
<td>284.6107</td>
<td>130</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>1011</td>
<td>284.6086</td>
<td>144</td>
<td>1</td>
</tr>
</tbody>
</table>

To divide the data into training and test samples, the "sklearn.model_selection.train_test_split" function from the sklearn library is used.

Train_test_split function can take many different parameters [9,10]. In this paper, the technique accepts the arguments described below:
- x is a matrix of features;
- y is the response vector;
- test_size: this is the ratio of the test data to the specified data. For example, if with 100 records in the dataset, test_size = 0.1, then the test sample will contain 10% of all records, and the training sample will contain 90%, respectively [16].

In this paper, the technique takes the following values: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1)

The training set contains 183393 records, and the test set contains 20378 records (Figure 2).

```python
print(x_train.shape)
pdنت
print(x_test.shape)

(183393, 11)
(20378, 11)
```

Fig. 2. The number of records in the training and test samples [Compiled by the authors].

When analyzing the requirements for the system being developed, it was decided to use a feedforward neural network to predict AQI [6,11]. To perform this task, the Python language and the tensorflow.keras library were chosen to implement work with a neural network.

The developed forecasting model is based on the following data sets: suspended particles (PM2.5, PM10), ozone (O3), nitrogen oxide (NO2), sulfur oxide (SO2) and carbon monoxide (CO), wind direction, wind speed, air temperature, weather conditions (snow, rain, etc.), cloudiness, pressure [12,13,14].

At the output of the neural network, we get one index, that is AQI for the current situation.
11 indicators and an annual data set are submitted to ANN input. It was experimentally found out that there would be a three-layer feedforward network to solve the problem.

The first two layers in the neural network are fully connected (Dense). ReLU function is used as the activation function [7,15]. The last layer is a response layer with Softmax activation function that returns the value from 1 to 10, which represent our AQI.

The first layer of the neural network consists of 13 neurons, the hidden layer contains 128 neurons, the last one contains 10 neurons (Figure 3)

```
model = Sequential()
model.add(Dense(13, activation='relu'))
model.add(Dense(128, activation='relu'))
model.add(Dense(10, activation='softmax'))
```

**Fig. 3.** Network structure in Python [Compiled by the authors].

As a result of training the neural network with the data set that was formed, its forecasting power on the test sample was 98.2% (Figure 4).

```
print(model.evaluate(x_test, y_test))
```

```
636/636 [============================] - 1s 1ms/step - loss: 0.0796 - accuracy: 0.9816
```

**Fig. 4.** Result of neural network training [Compiled by the authors].

A comparison was made between the real AQI results and those forecasted by the neural network (Figure 5).

To develop a feedforward neural network in C#, the Neuron class was used which is the main component of a neural network. The neuron class describes the activation function and the value of the neuron itself.

To connect neurons and set weights Synapse class was implemented.

The class of the neural network itself (Neural Network), which describes the input layer, hidden layers, output layers, as well as the interrelation of neurons in the layers.

**Fig. 5.** Comparison of forecasting and real AQI results [Compiled by the authors].
Also in this class, the following techniques are implemented:
- normalization of input values;
- connecting neurons in a neural network;
- neural network training.

Neural network training class (TrainingNetwork). It implements the technique of error back propagation, balancing weights and calculating the mean-root-square error.

4 Discussion

As a result of neural training, its forecasting power by one example is 69.41% (Figure 6).

Fig. 6. Result of neural network training [Compiled by the authors].

5 Conclusion

This work, results in the development of a valuable system for analyzing and forecasting the probability of air pollution by emissions from enterprises.

After implementing ergonomic systems, we can compare the accuracy of our systems, as in Python the accuracy is about 98%, and in C# it is a little less than 70%. So, it brings to the conclusion that the neural network in Python has learned better, most likely this is due to the mathematical functions that are used in Python library called Keras. In its turn, the neural network in C# learned much faster. Also, in the ergonomic system in Python there are many more methods to work.

References


13. R. Sangeetha, T.G. Ramabharathi, Test Engineering and Management 83(5), 6140-6146 (2020)

