World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:13, No:12, 2019

A Convolution Neural Network PM-10 Prediction System Based on a Dense Measurement Sensor Network in Poland

Authors : Piotr A. Kowalski, Kasper Sapala, Wiktor Warchalowski

Abstract : PM10 is a suspended dust that primarily has a negative effect on the respiratory system. PM10 is responsible for attacks of coughing and wheezing, asthma or acute, violent bronchitis. Indirectly, PM10 also negatively affects the rest of the body, including increasing the risk of heart attack and stroke. Unfortunately, Poland is a country that cannot boast of good air quality, in particular, due to large PM concentration levels. Therefore, based on the dense network of Airly sensors, it was decided to deal with the problem of prediction of suspended particulate matter concentration. Due to the very complicated nature of this issue, the Machine Learning approach was used. For this purpose, Convolution Neural Network (CNN) neural networks have been adopted, these currently being the leading information processing methods in the field of computational intelligence. The aim of this research is to show the influence of particular CNN network parameters on the quality of the obtained forecast. The forecast itself is made on the basis of parameters measured by Airly sensors and is carried out for the subsequent day, hour after hour. The evaluation of learning process for the investigated models was mostly based upon the mean square error criterion; however, during the model validation, a number of other methods of quantitative evaluation were taken into account. The presented model of pollution prediction has been verified by way of real weather and air pollution data taken from the Airly sensor network. The dense and distributed network of Airly measurement devices enables access to current and archival data on air pollution, temperature, suspended particulate matter PM1.0, PM2.5, and PM10, CAQI levels, as well as atmospheric pressure and air humidity. In this investigation, PM2.5, and PM10, temperature and wind information, as well as external forecasts of temperature and wind for next 24h served as inputted data. Due to the specificity of the CNN type network, this data is transformed into tensors and then processed. This network consists of an input layer, an output layer, and many hidden layers. In the hidden layers, convolutional and pooling operations are performed. The output of this system is a vector containing 24 elements that contain prediction of PM10 concentration for the upcoming 24 hour period. Over 1000 models based on CNN methodology were tested during the study. During the research, several were selected out that give the best results, and then a comparison was made with the other models based on linear regression. The numerical tests carried out fully confirmed the positive properties of the presented method. These were carried out using real ‘big’ data. Models based on the CNN technique allow prediction of PM10 dust concentration with a much smaller mean square error than currently used methods based on linear regression. What's more, the use of neural networks increased Pearson's correlation coefficient (R^2) by about 5 percent compared to the linear model. During the simulation, the R^2 coefficient was 0.92, 0.76, 0.75, 0.73, and 0.73 for 1st, 6th, 12th, 18th, and 24th hour of prediction respectively.

Keywords : air pollution prediction (forecasting), machine learning, regression task, convolution neural networks
Conference Title : ICAQPM 2019 : International Conference on Air Quality, Pollution and Management
Conference Location : Kuala Lumpur, Malaysia
Conference Dates : December 05-06, 2019