Interpolation and Prediction of $PM_{2.5}$ based on Conditional Generative Adversarial Network and a forecasting model

Abstract

Currently, air pollution is a severe problem, because pollutants such as Particulate matter of 2.5 micrometers affect human health. Therefore, several works address the prediction of this pollutant, using statistical methods and machine learning. However, these predictions are performed in places of a city, where air quality monitoring stations are available, which is not always possible due to their high implementation and maintenance costs. Thus, in this work, we propose an architecture based on a Conditional Generative Adversary Network to create new synthetic data and interpolate this pollutant in places where monitoring stations are missing.

1 Introduction

Air pollution is an essential problem at this time. One well-known pollutant is the Particulate Matter of 2.5 micrometers ($PM_{2.5}$); this is dangerous for human health because it causes respiratory and cardiovascular problems [1].

Recent works on this topic that addresses the prediction of this pollutant are carried out only in some places of a city, where air quality monitoring stations are available.

Therefore, to solve this limitation we propose an architecture based on a Conditional Generative Adversarial Network (cGAN) [2], from its capacity of creating synthetic data to interpolate this pollutant in places where air quality monitoring stations are missing. Moreover, we will perform a comparative study of forecasting models to find the best prediction of this pollutant, considering real and synthetic data.

2 Related work

There are two methods to the emission, dispersion, and prediction of pollutant concentrations, which are deterministic methods and statistic methods. Deterministic Methods adapt meteorological principles based on Atmospheric physics and chemical models to simulate the diffusion and dispersion of pollutant concentrations in a region-specific, two well-know models of this type are WRF-Chem [3] and CMAQ [4].

On the other hand, statistic methods generally are based on ARIMA [5], Support Vector Regression (SVR) [6] and Support Vector Machine (SVM) [7]. In addition, conventional neural networks are also generally considered, such as; MLR [8], RBF NN [9]. Deep neural networks, such as; Long-term memory and its variations, such as LSTM [10] and LSTMED [11], and the combination of a stacked automatic encoder (SAE) and a logistic regression (LR) [12]. About interpolation, Deep Air Learning (DAL) [13] is an architecture based on space-time learning to interpolate pollutants.

3 Proposal and current progress

We propose an architecture based on cGAN and forecasting model to interpolate and predict $PM_{2.5}$. We got a dataset of Beijing, which contained the information of meteorological variables and pollutant values from 35 air quality monitoring stations distributed in Beijing, then we considered $PM_{2.5}$ values and meteorological data such as real data for cGAN, to generate new synthetic values. After, we will process real data and synthetic data by forecasting models. See Figure 1.

Currently, the interpolation part is done, we selected data from 34 air quality monitoring stations as training set and one monitoring station as testing set, where we used fully connected layers in generator and discriminator, obtaining a root mean square error (RMSE) of 0.00589168 in the testing, in figure 2 we can see the comparison between the actual data of the testing set and the data interpolated by cGAN.

Therefore, in the Figures 3a and 3b we can see the result of the interpolation to all places, without monitoring stations, considering two date-times with one hour of differential.

Figure 1: Architecture to interpolation and prediction of $PM_{2.5}$. Source: Own.

Figure 2: Scatter plot. Comparing real $PM_{2.5}$ with interpolated $PM_{2.5}$. Source: Own.

Figure 3: Two interpolations of $PM_{2.5}$ in two different date-times. Source: Own.
References


[9] Bin Zou, Min Wang, Neng Wan, J Gaines Wilson, Xin Fang, and Yuqi Tang. Spatial modeling of pm2.5 concentrations with a multifactoral radial basis function neural network. Environmental Science and Pollution Research, 22, 03 2015.


