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

Anonymous Author(s)
Affiliation
Address
email

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 LSTME [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.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

3 Proposal and current progress

We propose an architecture based on cGAN and forecasting model to interpolate and predict $PM2.5$. 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.

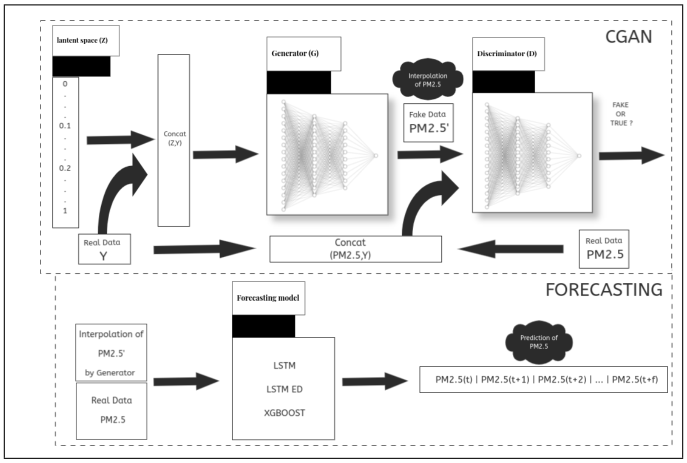

Figure 1: Architecture to interpolation and prediction of $PM_{2.5}$. Source: Own.

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.

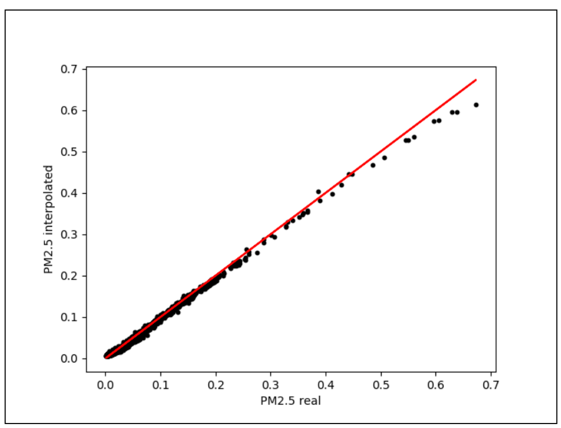

Figure 2: Scatter plot. Comparing real $PM_{2.5}$ with interpolated $PM_{2.5}$. Source: Own.

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.

(a) 2017-04-03 03:00:00

(b) 2017-04-03 04:00:00

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

References

[1] Environmental Protection Agency. Integrated science assessment for particulate matter. Technical Report EPA/600/R-08/139F, U.S. Environmental Protection Agency, jan 2009.

[2] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *ArXiv*, abs/1411.1784, 2014.

[3] Pablo E. Saide, Gregory R. Carmichael, Scott N. Spak, Laura Gallardo, Axel E. Osses, Marcelo A. Mena-Carrasco, and Mariusz Pagowski. Forecasting urban pm10 and pm2.5 pollution episodes in very stable nocturnal conditions and complex terrain using wrf–chem co tracer model. *Atmospheric Environment*, 45(16):2769 – 2780, 2011.

[4] Jianjun Chen, Jin Lu, Jeremy C. Avise, John A. DaMassa, Michael J. Kleeman, and Ajith P. Kaduwela. Seasonal modeling of pm2.5 in california's san joaquin valley. *Atmospheric Environment*, 92:182 – 190, 2014.

[5] W. Wang and Y. Guo. Air pollution pm2.5 data analysis in los angeles long beach with seasonal arima model. In *2009 International Conference on Energy and Environment Technology*, volume 3, pages 7–10, Oct 2009.

[6] P.J. García Nieto, E.F. Combarro, J.J. del Coz Díaz, and E. Montañés. A svm-based regression model to study the air quality at local scale in oviedo urban area (northern spain): A case study. *Applied Mathematics and Computation*, 219(17):8923 – 8937, 2013.

[7] Stanislaw Osowski and Konrad Garanty. Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Applications of Artificial Intelligence*, 20(6):745 – 755, 2007.

[8] Anastasia K. Paschalidou, Spyridon Karakitsios, Savvas Kleanthous, and Pavlos A. Kassomenos. Forecasting hourly pm10 concentration in cyprus through artificial neural networks and multiple regression models: implications to local environmental management. *Environmental Science and Pollution Research*, 18(2):316–327, Feb 2011.

[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.

[10] Xiang Li, Ling Peng, Xiaojing Yao, Shaolong Cui, Yuan Hu, Chengzeng You, and Tianhe Chi. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, 231:997 – 1004, 2017.

[11] Vikram Narasimha Reddy and Shrestha Mohanty. Deep air : Forecasting air pollution in beijing , china. 2017.

[12] Xiang Li, Ling Peng, Yuan Hu, Jing Shao, and Tianhe Chi. Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 23:22408–22417, 2016.

[13] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang. Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowledge and Data Engineering*, 30(12):2285–2297, Dec 2018.