Ozone Concentration Prediction

Lijing Zhai, Minyu Ma, Zephren Collison, Zongrun Li

Introduction

Ozone is an air pollutant which can cause harmful respiratory and other health effects on humans [1]. Predicting ozone concentrations in different conditions is useful for ozone control and public health management. Ozone concentrations depend on both weather conditions (wind, temperature, etc.) and emissions; especially VOC (volatile organic carbon) and NO_x (NO_2 + NO). These interactions can be empirically modeled using machine learning models trained by observation data to predict ozone levels over time [2]. Furthermore, multiple models can be utilized to predict ozone concentration between weather observation stations or even to forecast ozone levels under different weather and emission conditions.

Problem Definition

We will focus on the South Coast Air Basin (SoCAB) region in California from 1980 to 2020. Since myriad variables potentially affect ozone concentration, we will identify which variables are more important for predicting. Also, relative influence of these identified variables on ozone levels will be studied using different machine learning models.

Methods

The ozone concentration and weather data are provided by the EPA. A weather simulation model based upon satellite data is another potential data source if weather data is too limited. Emissions of VOC and NO_x are estimated by previous work [2]. Also some geographic information such as elevation, latitude, longitude, and time will be included since these affect solar flux; which is an important driving force for chemical reactions in the atmosphere. In this project, we will provide two routines to train the empirical models. First we will use principal component analysis to reduce the dimensions of our data and then use polynomial regression on the reduced data to predict ozone concentrations. The other method we will use is a self-organizing map to reduce dimensions and then an artificial neural network to predict ozone levels [3]. We can potentially compare these two model training routines by using all features directly. Random forest is another potential algorithm since it has better performance for discrete relationships than the supervised algorithms we listed above [4]. To avoid overfitting and estimate the uncertainty of our model, data withholding will be conducted.

Potential Results and Discussion

We will build 2-3 different models based on several machine learning algorithms and evaluate them on accuracy and robustness. Since previous work shows pollutants can be predicted by empirical data [2], we believe that the ozone levels can be predicted by machine learning methods based on observations.

If the ozone concentration can be estimated by our empirical models, there are several potential applications for the models. First is that spatial distribution of ozone can be predicted by applying the model for all places in SoCAB. The ozone levels across the entire region can be used for estimating ozone exposure (the product of ozone concentration and population). Ozone exposure is a helpful metric for medical research related to detrimental health effects of ozone [5]. Another potential application is the simulation of ozone levels under different emission conditions. These simulation results can be used for environmental engineers to help decide on emission control policies.

Reference

[1] Balmes, J. R. (1993). The role of ozone exposure in the epidemiology of asthma. *Environmental health perspectives*, *101*(suppl 4), 219-224.

[2] Qian, Y., Henneman, L. R., Mulholland, J. A., & Russell, A. G. (2019). Empirical development of ozone isopleths: Applications to Los Angeles. *Environmental Science & Technology Letters*, *6*(5), 294-299.

[3] Andras, P. (2013). Function approximation using combined unsupervised and supervised learning. *IEEE transactions on neural networks and learning systems*, *25*(3), 495-505.

[4] Friedman, J. H. (1994). An overview of predictive learning and function approximation. *From statistics to neural networks*, 1-61.

[5] Jerrett, M., Burnett, R. T., Pope III, C. A., Ito, K., Thurston, G., Krewski, D., ... & Thun, M. (2009). Long-term ozone exposure and mortality. *New England Journal of Medicine*, *360*(11), 1085-1095.