Predicting CO and NOx Emissions from Gas Turbines with Various Machine Learning Models

Team members: Zihao Chen, Keqin Wang

24787 Fall 2021 Project Proposal - TA Advisor: Jadlhir Trivedi

Motivation and Introduction

When we were figuring out our project ideas, we were aiming to do something meaningful and we found out we were both interested in the energy sector. How can machine learning help us to achieve a more environmentally friendly world? As engineers, what can we do with the skills and toolkits we have to predict emissions from a certain type of generator at the design phase? With these questions in mind, we explicitly searched for datasets that analyses energy generation emissions.

We anchored this dataset that was collected by Heysem Kaya, Pınar Tufekcl, and Erdinç Uzun aiming to understand attributes of a gas turbine generator and the corresponding CO and NOx emissions from the generator[1]. The detailed dataset will be described in later sections. Our goal is to build our machine learning model and compare our performance with the algorithm proposed in the paper, and a couple of other famous algorithms like random forest, logistic regression, etc. We will work together and split the work equally throughout the project.

Data Description

As one of the biggest open-source machine learning repositories, UCI provides thousands of datasets, which is useful for us graduate students. We retrieved our dataset from the UCI repository. We are using the 'Gas Turbine CO and NOx Emission Data Set Data Set'[1] by Heysem Kaya, Pınar Tufekcl, and Erdinç Uzun. Kaya et al. used the following attributes, Ambient temperature (AT), Ambient pressure (AP), Ambient humidity (AH), Air filter difference pressure (AFDP), Gas turbine exhaust pressure (GTEP), Turbine inlet temperature, Turbine after temperature (TAT), Compressor discharge pressure (CDP), and Turbine energy yield (TEY), to predict the generator's CO and NOx emissions. The dataset has 36733 instances and was arranged in a chronicle order so that it reflects a continuous five years of real data from the working generator. Kaya et al. used the first three years of data to train the model and tested the model on the latter two years of data. We propose to follow the same routine to better show the comparison between different algorithms.

Proposed Work

Machine Learning algorithms could be divided into three types: supervised learning, unsupervised learning, and reinforcement learning. We will be focusing on supervised learning because of the properties of our dataset. Here are steps we are proposing throughout this project:

- Given the chronicle order and the completeness of the dataset, we are replicating the routine of using the first three years' data to train our models and test the models on the latter two years' data.
- We will be comparing our models with various models that are common or will be covered in-class, including but not limited to basic regression models like linear regression, SVM, supervised learning models like KNN, and other models we can think of along the way.
- We will explore to strengthen our model through feature engineering if required.
- Finally, we will compare the performance of different models mentioned above and tabulate our results. We aim to outperform the model used in the original paper due to the fact that the model used in the paper was a one hidden layer feedforward network. We have the confidence to find a better model, with higher accuracy.
- Using our best model in our result, we hope we can generalize our model to a wider variety of generators with similar resolution in their attributes therefore we can also predict emissions from different types of generators.

Evaluation

For each developed model, we will evaluate the sum of square error (SSE) on the test dataset to determine the performance of the model. Lower SSE means that the model fits the data better. We will do a 10-fold run for each model, and compute their mean SSE as their final accuracy.

We are going to use the cross-validation to ensure the robustness of our study. We will emphasize on the comparison between our best model and the model developed in the paper followed by reasoning behind our model.

Reference

[1] Heysem Kaya, Pınar Tufekcl, and Erdinç Uzun. 'Predicting CO and NOx emissions from gas turbines: novel data and a benchmark PEMS', Turkish Journal of Electrical Engineering & Computer Sciences, vol. 27, 2019, pp. 4783-4796