

- What problem did you select and why did you select it?

The story behind the dataset is how to apply LSTM architecture to understand and apply multiple variables together. We choose this problem because we want to study on how to increase model accuracy towards forecasting.

- What database/dataset will you use? Is it large enough to train a deep network?

We will use a dataset that reports on the weather and the level of pollution each hour for five years at the US embassy in Beijing, China. The data includes the date-time, the pollution called PM2.5 concentration, and the weather information including dew point, temperature, pressure, wind direction, wind speed and the cumulative number of hours of snow and rain. It has over 43000 entries.

- What deep network will you use? Will it be a standard form of the network, or will you have to customize it?

LSTM. We will use a standard form of the network.

- What framework will you use to implement the network? Why?

TensorFlow: it provides extensive, flexible features, and an exhaustive library.

- What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?

We will read some paper about LSTM and air pollution, such as

<https://paperswithcode.com/paper/adversarially-trained-lstms-on-reduced-order>

<https://paperswithcode.com/paper/adversarial-autoencoders-and-adversarial-lstm>

We will also combine what we learn about LSTM in course to make improvement.

- How will you judge the performance of the network? What metrics will you use?

In this project, we will use root mean squared error to judge the performance of the network. It corresponds to the square root of the average of the squared difference between the target value and the value predicted by the regression model.

- Provide a rough schedule for completing the project.

Week 1. Search for dataset. Look for papers related to LSTM model and our research topic.

Week 2. Clean up dataset.

Week 3-4. Construct network framework.

Week 5. Fine tune network. Evaluate metrics to increase model accuracy.

Week 6. Wrap up for report and presentation.