A Deep Learning Framework for Air Quality Prediction

Group Nr. 13
Vicent Descals Carbonell, Priya Prabhakar

28th October 2022

1 Introduction and summary of the selected paper

- The general topic of research is to study spatial-temporal data in order to make predictions. In this case we are going to study air quality levels in different cities in India.
- The main idea on the original paper is to model spatial dependencies and temporal dynamics as the key of traffic predictions. [1]
- In the original paper implemented a Spatial-Temporal Dynamic Network to make accurate traffic predictions for real-world applications. The use two real-world datasets, NYC-Taxi dataset (40 days for training and 20 days for testing) and the NYC-Bike dataset (40 days for training and 20 days for testing).

2 Problem statement

With the development of industry, air pollution has become a serious problem. Monitoring it and understanding its quality is of immense importance to our well-being. Prediction is important for Current India Air Quality Index (AQI) is 189 POOR level with real-time air pollution.
The problem statement is to use deep learning framework as proposed in [1] to estimate air quality in India.

3 Research questions

- How to implement Air Quality prediction in India?
- How to handle spatial similarity and temporal periodic similarity jointly in Air Quality data?

4 Methodology

In order to implement a system to make accurate Air Quality predictions we will start looking carefully to the data we have and select the variables that are more relevant to this task. After that, we will use the STDN system developed in the original paper [1] as a guidance and we will adapt it to perform with our data.
In this project we will have to deal with spatial-temporal data, so we will need to use Auto-regressive models to work with our data and make predictions.

5 Evaluation approach

- **Metrics:** Given that we face a similar problem to the one presented in the original paper, we are going to use the same evaluation metrics. These are Mean Average Percentage Error (MAPE) and Rooted Mean Square Error (RMSE).
- **Baselines:**

  We will compare our RMSE and MAPE results to the ones presented in this paper [2].

6 Data sources and other resources

- Air quality data (daily) for 26 cities in India from 2015-2020.
  https://www.kaggle.com/rohanrao/air-quality-data-in-india
  https://cpcb.nic.in/
- Code and data of reference paper [1]
  https://github.com/tangxianfeng/STDN

Ethical statement

The work that will be done is this project has no ethical considerations. It is a study of the air quality in India and it's aim is to be able to predict accurately future values for a specific time and location. There are no privacy concerns as this data has been made publicly available by the Central Pollution Control Board for everyone.
Our work cannot be use in a harmful way, but it can be used by companies and governments to know which are the more affected areas and create solutions to reduce pollution and improve the quality air in those areas.

Division of workload

For this project we will do every task together because it's very difficult to clearly divide the task. At a high level these are the tasks we are going to go through it:

- Reproduce results from the original paper.
- Adapt the original code to make it work with the new dataset.
- Run the experiments.
- Compare results with the other paper mentioned above.
- Write the final report.

Code

- Link to the Github repository:
  https://github.com/videscar/Urban-Computing-Project

References

[1] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng and Zhenhui Li. "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 5668–5675. URL: https://arxiv.org/pdf/1803.01254.pdf.

[2] Gaganjot Kaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu and Gang Xie. "Air quality prediction: Big data and machine learning approaches". In: *Int. J. Environ. Sci. Dev* 9.1 (2018), pp. 8–16.