Generative Adversarial Network for Temporal Window imputation for Multivariate Time Series Data

Problem Description

Missing data is a common challenge that prevents from performing useful data analysis tasks. In particular, in the IoT domain, this can be due to broken or faulty sensors that fail to record certain signals for an unknown and unpredictable amount of time.

Thus, in order to still be able to extract value from the data collected to make useful predictions, it is important to impute missing values.

We suggest the problem of data imputation for the Telenor Air Quality prediction IoT use case, where the data consists of measurements of the concentration of certain air pollutants in Trondheim having the form of multivariate time series. The goal is to impute missing temporal windows, by leveraging on correlations between the components of such multivariate time series.

An interesting method that could be explored consists of Generative Adversarial Networks, which have proven to perform exceptionally well in the data imputation problem for non time series datasets [1]. An adaptation of GANs to the time series domain was developed in [2,3] but so far, no research on temporal window imputation has been performed yet. Stating the problem in these terms would change the perspective to reconstructing the partial missing signal by means of the complex underlying correlations among the different time series.

Data

Trondheim Air Quality dataset and in alternative other open datasets can also be considered, we recommend the ones used for testing in [2,3].

Task

The challenge of this project would be to develop a novel temporal window imputation method for multivariate time series. This could be carried out in the following way:

- Produce a literature review on SOTA data imputation methods with a focus on GANs as they are particularly well suited for this task.
- Test existing data imputation methods and extend them to window imputation. The testing is recommended to be performed using different datasets, together with the Air Quality dataset provided by Telenor, in order to evaluate performance and develop an algorithm that is more robust.
- Develop a novel window imputation method based on the algorithms explored above.

References:

[1] Yoon, Jinsung, James Jordon, and Mihaela Schaar. "GAIN: Missing Data Imputation using Generative Adversarial Nets." *International Conference on Machine Learning*. 2018.

[2] Luo, Yonghong, et al. "Multivariate time series imputation with generative adversarial networks." *Advances in Neural Information Processing Systems*. 2018.

[3] Luo, Yonghong, et al. "E2GAN: end-to-end generative adversarial network for multivariate time series imputation." *Proceedings of the 28th International Joint Conference on Artificial Intelligence.* AAAI Press, 2019.

Academic supervisor: **Massimiliano Ruocco**, Associate Professor II at NTNU, and Senior Research Scientist at Telenor Research.
(massimiliano.ruocco@ntnu.no)
Co-supervisor: **Sara Malacarne**, Research Scientist at Telenor Research.
(sara.malacarne@telenor.com)