Research Project

Graph signals forecasting with application to transportation

Supervisors: Alejandro de Concha and Argyris Kalogeratos
`{de_la_concha_duarte,kalogeratos}@ens-paris-saclay.fr`
Institution: Machine Learning and Masssive Data Analysis (MLMDA) research group,
Centre Borelli, ENS Paris-Saclay.
Keywords: Graph processes, graph signal processing, time-series forecasting, transportation.

1. Context and motivation

A broad and diverse range of human and natural processes can be seen as time-evolving processes taking place over the nodes of a graph. Take for instance the spread of a virus or false news in a social network, measurements recorded by a network of weather stations, or train delays observed in a railway system. Intuitively, the state of the system at time t depends on what has happened in the past and nodes that are close in the graph tend to exhibit similar behavior. Modelling such interplay between the behavior of the nodes in time as part of the spread from one side, and the graph structure, from the other, is a surprisingly recent topic. In particular, in such a context there is a lack of methods aiming to forecast the value at a given node v at a given time t taking into account the full view of the problem. The goal of this internship is to make a step further in this direction, which is to design methods that incorporates both components and have good performance in practice. The main application of interest will be the prediction of train delays in the SNCF railway system. This end will be facilitated by the participation of the SNCF R&D department to the project, hence the student will have the opportunity to interact with both the academic and the industrial labs.

2. Scientific objectives

A graph process is a multivariate time series defined over an underlying graph structure. In the literature, there have been two main approaches to address the problem of foresting multivariate time series: Graphical Models (GM) and Graph Signal Processing (GSP) [1, 2]. In the GM approach, the behavior of a multivariate time series is understood in terms of its conditional dependencies encoded in a graph G. G should be inferred and, in the Gaussian case, this is equivalent to estimate the covariance matrix of the graph processes. On the other hand, under the GSP perspective, if the graph is already known, the interplay between the graph and the observations is understood via harmonic analysis techniques. Our intention, is to mainly focus in the second approach since we will suppose the graph structure to be known as in the case of transportation networks.

GSP is a relatively new field whose main contribution is the extension of classical signal processing techniques to signals observed over graphs (graph signals) [3, 4]. The warhorse of GSP is the concept of Graph Fourier Transform (GFT) which allows us to redefine the notion of frequency, filters and stationarity in the domain induced by the graph. In [5–7] a further step is taken, the Fourier Transform is extended to include the time and the vertex domain jointly. This technique has interesting applications. For instance, it is possible to find interpretable spectral Fourier representations of graph processes and to localize the source of a graph event such an earthquake [6,7]. Furthermore, in [2,8–10], it is shown how predictions can be improved when working with both domains at the same time.

In [9], the authors introduce an optimal sampling strategy to recover a graph processes from a given subset of nodes. The strategy performs better compared with previous works in the GSP literature where the time component is ignored. The authors of [10] introduce a change-point detection method based on the GFT aiming to segment a stationary graph process and recover a sparse representation of it with minimum intervention from the user. In [2, 8], a model called G-VARMA is introduced aiming to forecast the value at a given node v at time t, the same problem we are interested in. The idea is to translate the problem to the spectral domain with the GFT and, then, fit temporal ARMA models for each of the Graph Fourier Coefficients. This model outperforms others that do not consider the correlations induced by the graph. Nevertheless, the G-VARMA model requires the graph processes to be stationary in the temporal and graph domain. This hypothesis imposes strong restrictions in the covariance matrix of the graph processes that are hardly satisfied in practice and limits significantly its applicability to transportation networks.

Goal: The goal of the project is to design a method that can deal with not stationary graph processes. A suggested strategy is to model the time component in a similar way as in the state-of-the-art model proposed in [11] designed for univariate time series and exploit the GSP toolbox to extend it to signals observed over graphs.

3. Planning

The 6-month work plan is as follows. During the first one third of the time, the student will focus on reviewing the existing literature as well as the necessary data prepossessing to apply these methods to the SNCF dataset. In the rest of the internship, the student will experiment with different approaches in order to outperform a baseline model and collaborate with the SNCF research team to validate the pertinence of the suggested solutions.

References

[1] R. Dahlhaus and M. Eichler, “Causality and graphical models in time series analysis,” *Oxford Stat. Sci. Ser*, vol. 27, 01 2003.

[2] E. Isufi, A. Loukas, N. Perraudin, and G. Leus, “Forecasting time series with varma recursions on graphs,” *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4870–4885, 2019.

[3] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[4] A. Ortega, P. Frossard, J. Kovačević, J. Moura, and P. Vandergheynst, “Graph signal processing: Overview, challenges, and applications,” *Proc. of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[5] E. Isufi, “Graph-time signal processing: Filtering and sampling strategies,” 2019.

[6] A. Loukas and D. Foucard, “Frequency analysis of time-varying graph signals,” in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016, pp. 346–350.

[7] F. Grassi, A. Loukas, N. Perraudin, and B. Ricaud, “A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs,” *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 817–829, 2018.

[8] A. Loukas and N. Perraudin, “Predicting the evolution of stationary graph signals,” *CoRR*, vol. abs/1607.03313, 2016.

[9] E. Isufi, P. Banelli, P. D. Lorenzo, and G. Leus, “Observing and tracking bandlimited graph processes from sampled measurements,” *Signal Processing*, vol. 177, p. 107749, 2020.

[10] A. de la Concha, N. Vayatis, and A. Kalogeratos, “Offline detection of change-points in the mean for stationary graph signals,” 2020.

[11] S. Taylor and B. Letham, “Forecasting at scale,” *PeerJ Prepr.*, vol. 5, p. e3190, 2017.