Changepoint and anomaly detection in industrial monitoring with time series

Supervision:

- Thomas Moreau, MIND, Inria (thomas.moreau@inria.fr)
- Mathieu Barré, Saint Gobain Research (mathieu.barre@saint-gobain.com)

Context. Saint Gobain is a world leader in the habitat and construction markets. With more than 170,000 employees and a presence in 70 countries, Saint Gobain designs, produces, and distributes materials and solutions that are key ingredients in many domains: in buildings, transportation, infrastructure, and in many industrial applications. Real-time monitoring of production processes is a critical aspect of optimizing Saint-Gobain's activities, both in terms of cost and energy efficiency. In particular, more and more sensors are installed on production lines to monitor the evolution of the different parameters of these processes with maximum precision. Being able to use this data to ensure the proper functioning of the lines and to improve the production processes is a major challenge for the group.

Sensor data are time series that are often very noisy and contain a lot of information. They are used to ensure the quality of a process or even to control it. Their reliability and robustness are crucial, and it is very important to be able to detect abnormal situations or changes in regimes as soon as possible, as well as to classify them.

Methods. The goal of this internship is to study the problems commonly called anomaly detection on time series. Anomaly detection is a particular type of classification problem, where the goal is to detect abnormal situations in a dataset. By definition, these situations are rare, and no consistent annotations are available. The methodology often relies on characterizing the nominal behaviors of the systems and then detecting deviations from these behaviors. Starting from the available scientific literature, the intern will seek to develop a system for detecting abnormal situations and changes in regimes on multivariate time series from an industrial flat glass production process. To do so, a benchmark of existing methods will be carried out using `benchopt` [3], with open datasets and industrial data from Saint-Gobain production factories. The benchmark will draw inspiration from the recent Anomaly detection benchmark [1] and focus on comparing state-of-the-art methods for anomaly detection in time series such as `Kats`[1] and `TODS` [2]. If time allows it, the intern will also seek to consolidate these developments by contributing missing anomaly detection methods in these existing open-source libraries. The benchmark will also consider changepoint detection methods [4] as a particular type of anomaly for industrial processes. This literature review will serve as a starting point for the development of a new method adapted to the particular problem at hand.

Then, the candidate will seek to adapt the previous work on other applications of anomaly detection within the Saint-Gobain group. Depending on the progress and problems raised during the internship, it may lead to a thesis within Saint-Gobain Research Paris.

[1] https://github.com/facebookresearch/Kats

Environment. The internship will take place at Inria Saclay, in the MIND team for maximally 6 months (April-September 2024). This is a large team focused on mathematical methods for statistical modeling of brain function using neuroimaging data (fMRI, MEG, EEG). This team has strong expertise in time-series processing with machine-learning techniques and numerical optimization. The team also has a strong focus on scientific software development.

The intern will have strong interaction with the Saint-Gobain Datalab, an R&D unit composed of 20 data scientists at the service of all the group's activities. The group is composed of talented, curious, and friendly people, whose objective is to bring innovative and practical solutions using artificial intelligence to the different units of Saint-Gobain. In this context, the group works on various projects using different types of data collected by the production plants and aims at improving production with Industry 4.0 approaches.

Requirements. We seek candidates who are strongly motivated by challenging research topics in machine learning, time-series processing, and industrial process optimization. Applicants should have a strong mathematical background with knowledge of signal processing and machine learning. Basic knowledge of time-series analysis techniques would be a plus. With regards to software engineering, proficiency in `Python` is expected and preliminary experience in a deep-learning library is a plus.

References

[1] Songqiao Han, Xiyang Hu, Hailiang Huang, Minqi Jiang, and Yue Zhao. ADBench: Anomaly Detection Benchmark. *Advances in Neural Information Processing Systems*, 35:32142–32159, December 2022.

[2] Kwei-Herng Lai, Daochen Zha, Guanchu Wang, Junjie Xu, Yue Zhao, Devesh Kumar, Yile Chen, Purav Zumkhawaka, Minyang Wan, Diego Martinez, and Xia Hu. TODS: An Automated Time Series Outlier Detection System. In *AAAI Conference on Artificial Intelligence*, volume 35, pages 16060–16062, May 2021.

[3] Thomas Moreau, Mathurin Massias, Alexandre Gramfort, Pierre Ablin, Pierre-Antoine Bannier, Benjamin Charlier, Mathieu Dagréou, Tom Dupré la Tour, Ghislain Durif, Cassio F. Dantas, Quentin Klopfenstein, Johan Larsson, En Lai, Tanguy Lefort, Benoit Malézieux, Badr Moufad, Binh T. Nguyen, Alain Rakotomamonjy, Zaccharie Ramzi, Joseph Salmon, and Samuel Vaiter. Benchopt: Reproducible, efficient and collaborative optimization benchmarks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, New-Orleans, LA, USA, November 2022. Curran Associates, Inc.

[4] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, February 2020.