Data-driven predictive maintenance and time-series applications

Kefalas, M.

English Summary

Prognostics and health management (PHM) systems play a significant role in Industry 4.0. PHM aims to maximize the operational availability of assets, reduce maintenance costs and improve reliability and safety by monitoring and predicting (prognostics) the future state of the asset. As such, PHM functions as a decision support tool aiming to make timely and better-informed maintenance decisions. This allows a shift from traditional maintenance strategies, such as reactive maintenance (RM) and preventive maintenance (PM), to what is known as *predictive maintenance (PdM)*. PdM estimates when maintenance should occur and thus, increases safety and maximizes usability by avoiding premature maintenance.

An essential task in PHM and, thus, PdM is failure prognostics and specifically the estimation of an asset's remaining useful life (RUL). As its name suggests, the estimation of the RUL allows for managing *pending* equipment failure and grants sufficient lead-time so that the necessary decisions, logistics, personnel, equipment, and spare parts can be organized and deployed, thus minimizing both equipment downtime and repair costs. In general, there are three major classes of approaches that deal with the RUL estimation. Namely, model-based, data-driven, and hybrid methods. Model-based methods (or physics-based methods) rely on established mathematical/physical models of the asset and consequently require a thorough understanding of the asset's physics and processes. This can sometimes be prohibitively costly, in terms of time and money, especially for very complex systems. On the other hand, data-driven methods are relatively easier to develop as they do not call for (a lot of) expert or domain knowledge to develop the model, rendering them domain-agnostic and easily transferable between domains and because of the plethora of tools that have been and are being developed. However, they require large amounts of (maintenance) data, a prerequisite that is not always possible due to the nature of certain applications. Machine learning (ML) and deep learning (DL) commonly belong to this category. Lastly, hybrid (or fusion) methods attempt to leverage the advantages of the two previous methods while minimizing their limitations by combining (or fusing) model-based approaches and augmenting them with data-driven methods and vice versa.

The relative ease of data-driven methods has, in a sense, democratized the RUL estimation (and PdM thereof) due to their (mostly) domain-agnostic nature and broad applicability. Nevertheless, there are various challenges to consider in the data-driven domain. The sufficient pre-processing of the (raw) data, the selection of the learning algorithm, and the optimization

of their hyperparameters are some non-trivial questions that researchers and end-users need to address. In this view, we investigate the usage of automated machine learning (AutoML) for RUL estimation and, further, propose an automated framework for configuring, in an end-to-end fashion, RUL prediction models.

The estimation of the RUL (and any other prognostics measure for that matter) lies at the heart of PHM and PdM. However, determining the RUL is only part of the overall promise of PdM (albeit a pivotal one). As its name suggests, PdM uses prognostics for maintenance planning. Therefore, in principle, having the estimated RUL values, one can plan the maintenance of the assets through scheduling. Optimizing the maintenance schedule, however, is a further challenge that needs to be dealt with. We illustrate this through the flexible job-shop scheduling problem (FJSSP), as it resembles more closely dynamic, real-world environments where operations can be processed on different sets of machines, and we show how heuristics can support the optimization process.

Understanding the data generating process is another important topic in data-driven PdM. Uncovering the relationship between features in the data allows one, for example, to create models that imitate the actual process and monitor the deviation between the predicted and the observed data. In turn, this can be used as an early-warning tool for a fault or failure. We show how explainable AI can facilitate the understanding of the data-generating process through a case study from the aviation industry.

Finally, we show how a method originally developed for PdM in the automotive industry can lend itself to the medical domain, exhibiting the significance of knowledge transfer and how it can impact research between different scientific fields. We investigate this through a case study from the field of Neurology.