Fraunhofer PORTUGAL

FRAUNHOFER CENTER FOR ASSISTIVE INFORMATION AND COMMUNICATION SOLUTIONS – AICOS

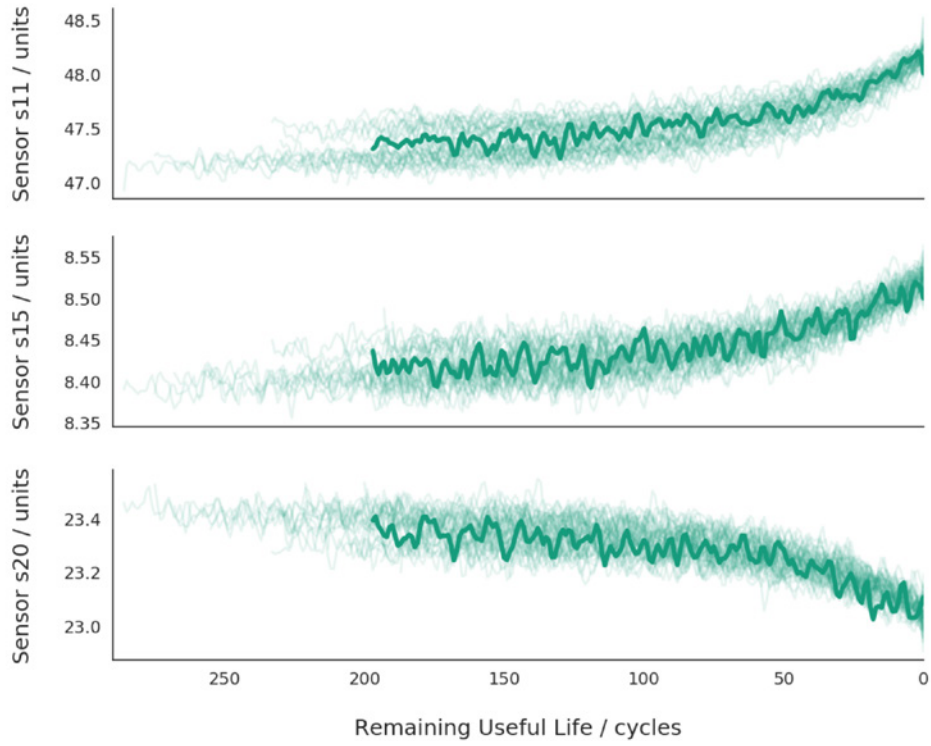
TSFEL: TIME SERIES FEATURE EXTRACTION LIBRARY

Keywords:
Time series, Machine learning, Feature extraction, Python

INTELLIGENT SYSTEMS

Authors:
Duarte Folgado
Marília Barandas
Hugo Gamboa

Over the last years, the technological breakthroughs motivated by the rise of Internet-of-Things led to the proliferation of sensors to measure a plethora of physical processes. Those observations often result in the creation of large quantities of data in the form of time series, which are described as sequences of numerical observations ordered in time. Large time series streams have hidden patterns that hold very relevant knowledge for the decision-making and policy planning of organisations. Forecasting the future based on historical pattern data is extensively employed in financial and business applications. Time series can also be used in industrial shop-floor management by providing insights to predictive maintenance by monitoring the asset's health and preventing downtime as illustrated in the following figure. The values measured by sensors help to diagnose whether a machine might be near breakdown, by predicting the remaining useful life based on historical measurements.



How to uncover the hidden knowledge of time series?

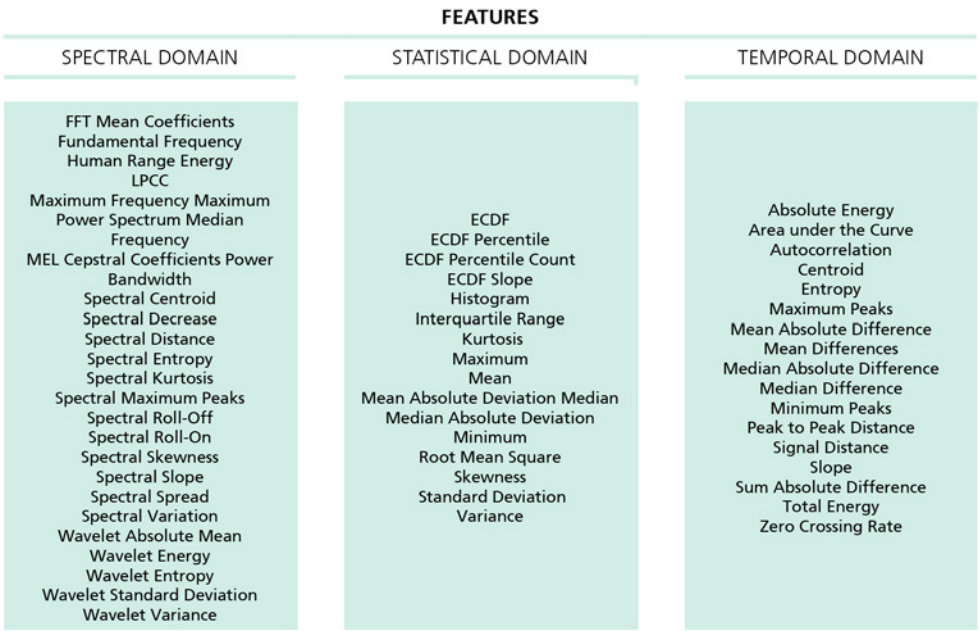
Analysing such a significant amount of data manually would be impracticable. This is where machine learning comes to the rescue, by automatically using automatic learners to digest and issue recommendations over large data streams.

The process of time series feature extraction is one of the preliminary steps in conventional machine learning pipelines and aims to extract a set of properties to characterise the temporal sequences. Feature extraction is an important step of the machine learning stack. However, it is often a time-consuming and complex task.

In a recent study titled "TSFEL: Time Series Feature Extraction Library", published in the journal SoftwareX, we propose a toolbox to support researchers for fast exploratory analysis supported by an automated process of feature extraction on multidimensional time series. TSFEL is open source and available at GitHub. This library is the result of an international collaboration between AICOS and the Cognitive System Lab, from the University of Bremen, Germany.

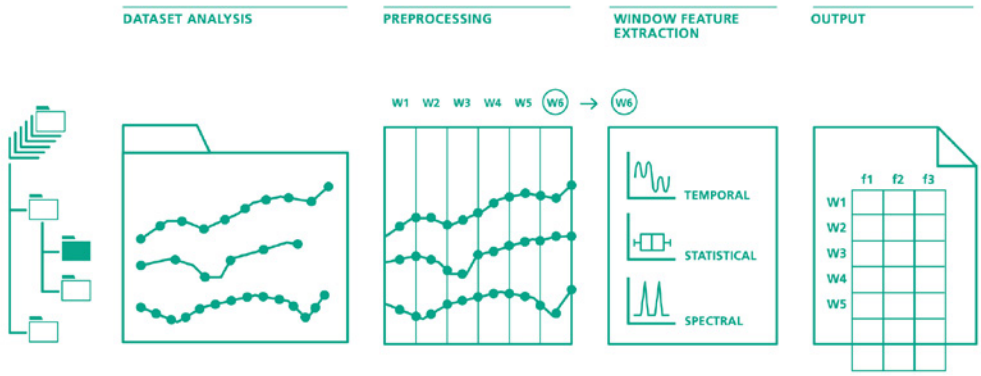
Time Series Features

Feature extraction is the process of computing a set of transformations to calculate features from raw data. Ideally, those features should encapsulate the properties of the dataset, by creating a lower-dimensional data representation that facilitates the learning process. TSFEL calculates over 60 different features in temporal, statistical, and spectral domains.



TSFEL is optimised for large-scale feature extraction on time series and the following figure illustrates the main steps of the TSFEL' pipeline. Time series are passed as inputs for the main TSFEL extraction method either as arrays previously loaded in memory or stored in files on a dataset. Since TSFEL can handle multidimensional time series, a set of preprocessing methods is applied afterwards to ensure that not only the signal quality is adequate, but also, time series synchronisation, so that the window calculation process is properly achieved. Time series are divided into windows, i.e., time intervals, from which the features are extracted. The result is saved using a standard

schema ready to be digested by most of the machine learning and data mining frameworks. Each line corresponds to a window, whereas the corresponding columns store the extracted features' values.



Users can interact with TSFEL in two forms: a backend built upon a Python package aimed at advanced users; a frontend displayed in online spreadsheets targeting beginners. For both cases, TSFEL also provides a crucial aspect for the deployment of machine learning algorithms in real scenarios – a comprehensive evaluation of the computational complexity of each feature.

Additionally, TSFEL also contains a collection of notebooks that can be instantiated with Google Colab, thus, requiring no installation whatsoever. Those notebooks can be used to learn methods for feature extraction, classification, and evaluation of machine learning techniques on time series data.

Computational Evaluation

All features implemented are classified by their time complexity according to the following known models: $\mathcal{O}(n^2)$, $\mathcal{O}(n \log n)$, $\mathcal{O}(n)$, $\mathcal{O}(\log n)$ and $\mathcal{O}(1)$. Firstly, a set of time series with incremental length is synthetically generated from a sinusoidal model. Secondly, the feature extraction runtime is calculated for each incremental length time series. This process allows creating a curve displaying the relationship between the time series length and the execution time for each feature. Thirdly, we use a non-linear least squares regression to fit the runtime results to the known models. For each feature, we assign the time complexity which minimises the χ^2 among all the time complexity models.

Impact on real use cases

With the proliferation of Cyber-Physical Systems, more data and processes are being digitised across several sectors. Whilst machine learning holds the promise to uncover hidden patterns of big data to gain new insights on data-driven businesses, the deployment of such systems might still be limited by scalability concerns. For example, in the context of large manufacturing plants, performing anomaly detection on their products based on sensors producing high data volume, requires a compromise between high accuracy and low latency. TSFEL helps to mitigate these risks by providing a comprehensive list of feature extraction methods on time series with an associated estimate of computational complexity, enabling us to have an idea of the computational cost of the feature extraction in the early stages of the machine learning stack. We are also currently researching how to run artificial intelligence algorithms on the Edge. With the increasing availability of powerful and low-energy consumption compu-

ting devices, Edge AI proposes a distributed computing paradigm that deploys models on the "edge" of the network, reducing bandwidth and latency and increasing security and decentralisation.

Future Directions

We are actively researching how to extract meaningful descriptors from time series data, by calculating and selecting the most relevant features towards an optimal accuracy/computational cost trade-off. TSFEL is the result of an effort to centralise the main time series feature extraction methods used by AICOS across different application scenarios. We hope this first step leads to further improvements through an active contribution of the research community.