FACULTY OF ENGINEERING AND ARCHITECTURE

IDLab INTERNET & DATA LAB

imec

GHENT UNIVERSITY

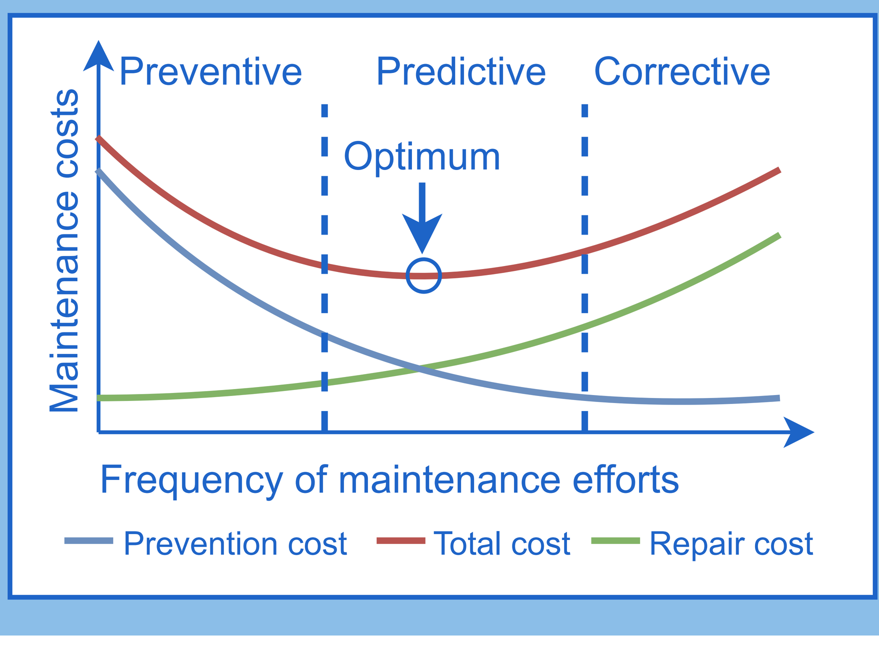
A semi-supervised anomaly detection approach detecting mechanical failures

Colin Soete, Michaël Rademaker and Sofie Van Hoecke / IDLab, Ghent University, imec, Ghent, 9000, Belgium

Problem statement

Breakdown of machinery is a costly expense for many manufacturing industries. Insufficient labels limiting them to unsupervised anomaly detection techniques often refrain companies for solving this issue with predictive maintenance. The lack of interpretability of machine learning models also leads to unjust reluctance of using these models.

Fig. 1 (right): Difference in maintenance costs regarding the various strategies of performing maintenance, such as preventive, predictive and corrective maintenance.



Use case

An appropriate example use case where a lack of labeled data is prevalent, and where maintenance costs escalate, is the pharmaceutical industry. The dataset contains a multivariate time series per machine, in this case several dryers and filters. However, this research is not limited to the presented use case but can also be deployed on other machines and/or use cases. The data in our setup are impacted by several different (latent) fault types which are mostly unique for each machine. The appliances are equipped with sensors that read out information such as temperature, pressure, and valve statuses.

Architecture overview

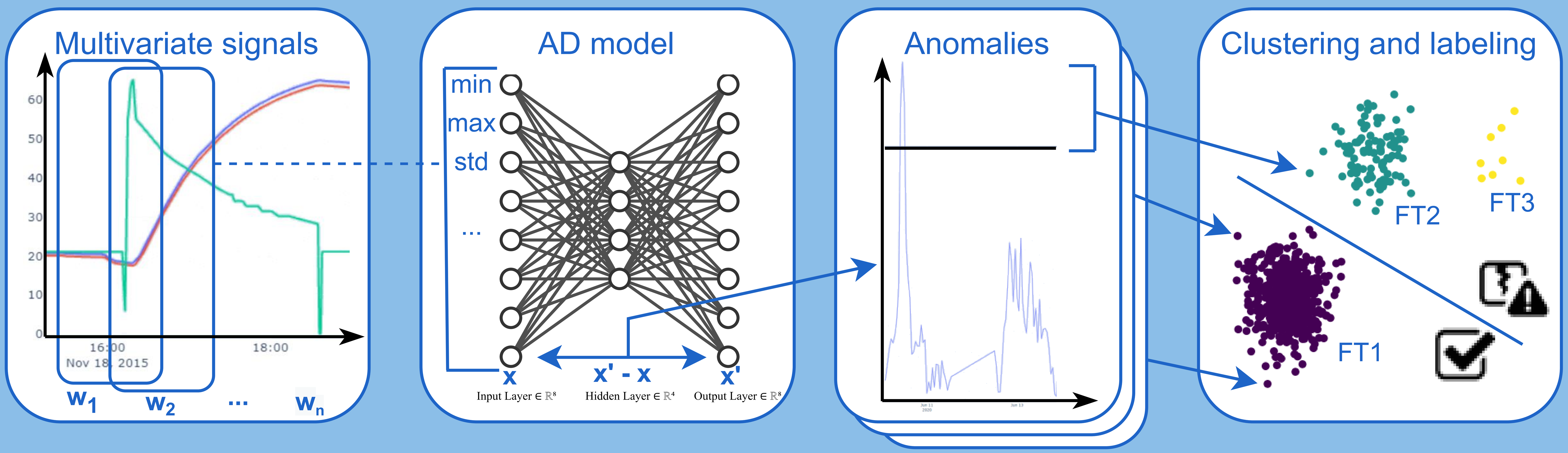


Fig. 2: The multivariate signals are segmented into windows, of which features are extracted. These features are then fetched into anomaly detection models which return anomalies. Next, the anomalies are transformed to a lower dimension space and clustered according to potential fault types. These fault types are then labeled according to the degree of damage. After labeling the clusters of similar fingerprints, a supervised model is learned that mimics experts' behaviour and identifies faults that show similar behaviour that has been seen before.

Results

Machine	# Samples	# Anomalies: Non-damaging	# Anomalies: Damaging	# Anomalies: Total	# Clusters	Labeling gain: $1 - \sum_{w/s=0}^{n} \frac{\#Clusters}{\#Anomalies_{w/s}}$
D001	15250	76	10	86	10	88.37 %
D009	18713	35	32	67	11	83.58 %
D020	5534	11	15	26	5	80.77 %
F005	28827	95	2	97	12	87.63 %
F008	26319	24	12	36	8	77.78 %
F012	24672	48	4	52	9	82.69 %
F017	26164	132	28	160	12	92.50 %

Table 1: General description on the number of samples, anomalies and clusters of all machines. The last column reports on the reduction of labeling resources, which mentions the labeling gain when compared to other labeling tasks. The formula represents the advantages gained that otherwise needed to be labeled in other research.

The transformation from an unsupervised system to a semi-supervised system was done by incorporating logistic regression models, which perform adequately, even with limited amount of training samples. Furthermore, the coefficients of these models give a measure of the most prominent features used to distinguish fault types from normal behaviour.

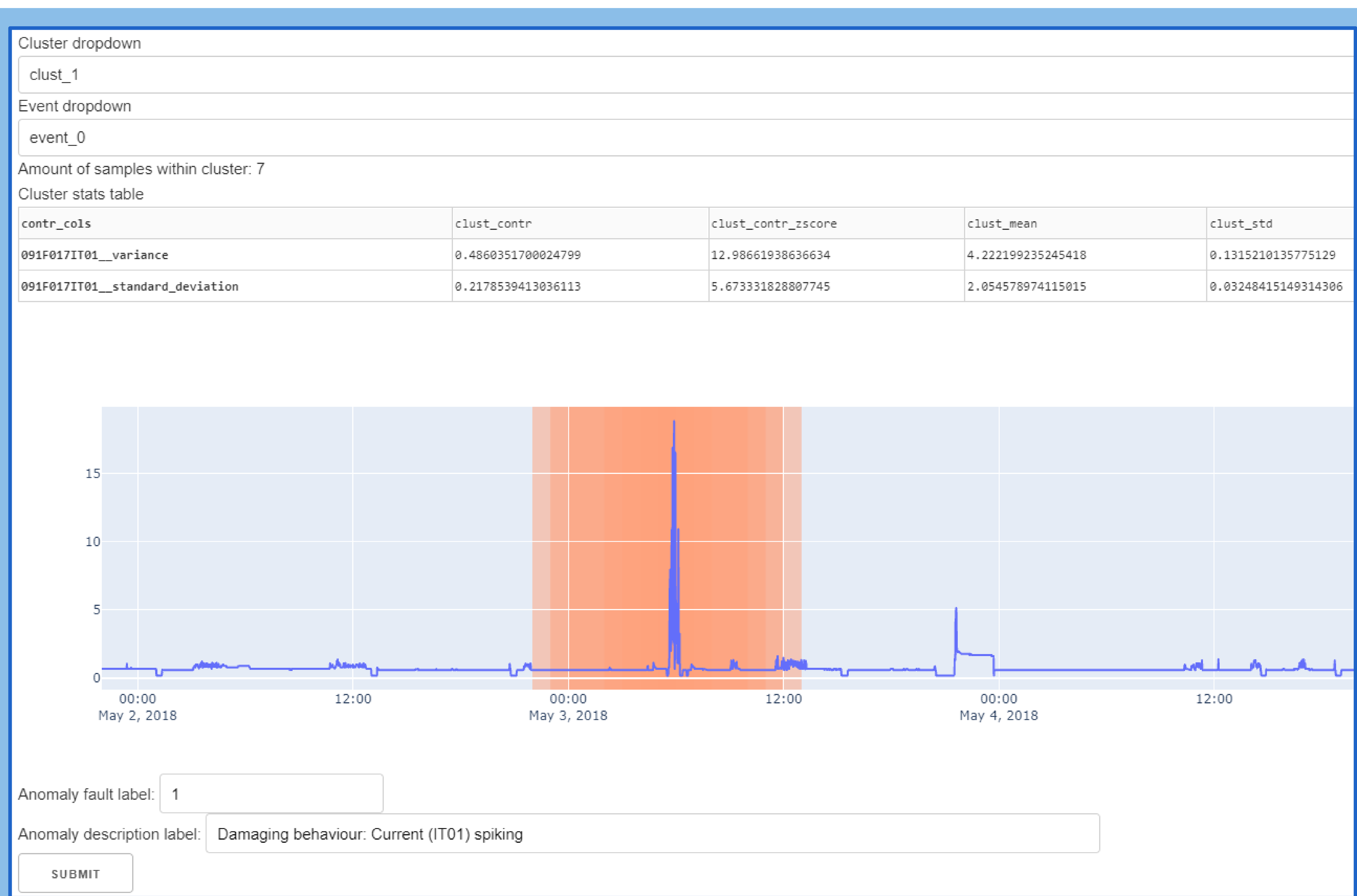


Fig. 3: Example of the labeling dashboard with an anomaly of filter F017. This cluster of anomalies is caused by a current sensor having odd variance and standard deviation values within the window(s) shown in orange. Hence, the cluster is labeled by an expert as damaging behaviour along with a small description.

Conclusions

- The results of this research look promising as the architecture is indeed able to recognize critical recurring faults with little labeling overhead, having an average labeling gain of 85% using our methodology.
- Thanks to our architecture, the resources needed to perform the labeling tasks were heavily reduced. By using a user-friendly dashboard, an expert could efficiently input his knowledge into the architecture.
- Ultimately, the labeled failure data allows creating better failure prediction models, which in turn enables more effective predictive maintenance resulting in increased profits.

Contact

colin.soete@ugent.be

colin-soete

0000-0002-2035-3466