Edge computing for advanced vibration signal processing

Timothy VERSTRAETEN[1]*, Felipe GOMEZ MARULANDA[1]*, Cédric PEETERS[2],
Pieter-Jan DAEMS[2], Ann NOWÉ[1], Jan HELSEN[2]

[1]Artificial Intelligence Lab Brussels
[2]Acoustics & Vibrations Research Group
Vrije Universiteit Brussel, Pleinlaan 2, 1050, Elsene

Abstract

Today, Industry 4.0 is being introduced. Machines are equipped with internet connection and increasingly sensorized using Industrial Internet of Things (IIoT) sensors. Especially the emergence of 5G is a game changer in this regard. It becomes possible to send data at high speeds to cloud computing data-centers. However, streaming all data is deemed to be unnecessary. It is more advantageous to use the additionally available bandwidth to drastically increase the number of connected sensors. Thus, on-board processing of the data directly at the edge is necessary. This paper illustrates this edge computing concept using data of wind turbines. Different fault indicators are calculated directly on an embedded system. In addition to statistical features also more complex signal processing pipelines combined with machine learning approaches are used. An example of a more advanced technique is the spectral coherence approach. This is one of the most promising approaches for bearing fault diagnostics to extract the optimal envelopes. This approach requires a significant amount of computational power. Today, different Advanced Risc Machine (ARM) processors are available in embedded architectures. Moreover, CPU based single board computers are available. Embedded GPUs allow dedicated machine learning algorithm processing. In this paper an NVIDIA Jetson device combining multiple ARM cores with a GPU is used. The edge computing concept is validated by processing pipelines on vibration and SCADA data originating from operational wind turbines using such architectures. Both healthy and faulty data sets are processed.

1 Introduction

In the context of Industry 4.0 efforts there is a continuously decreasing cost for sensors. As such the range of machines and other systems that are equipped with on-board instrumentation has increased substantially and will increase even more in the years to come. For those industries where the cost of downtime is high there is a strong interest and economic opportunity to move towards predictive maintenance. Therefore, more and more companies show interest to acquiring more data from their product for condition monitoring and design validation purposes. Continuous data collection allows to gain insights in product usage and thus forms the basis for design improvements from better understanding asset behavior in the field. Adequate processing algorithms are needed to perform usage evaluation and failure prediction to extract useful information from these sensors. Typically these algorithms use acceleration or current signals sampled at high frequency. The wide adoption of the Internet has brought broad coverage and continuous data connections at many locations all over the world. However, for many industrial applications the local connectivity can still be problematic due the limited bandwidth of wired or mobile connections. As such, streaming high frequency data is still unfeasible. Local processing is thus necessary and will become more important with increasing data volumes.

*These authors contributed equally.

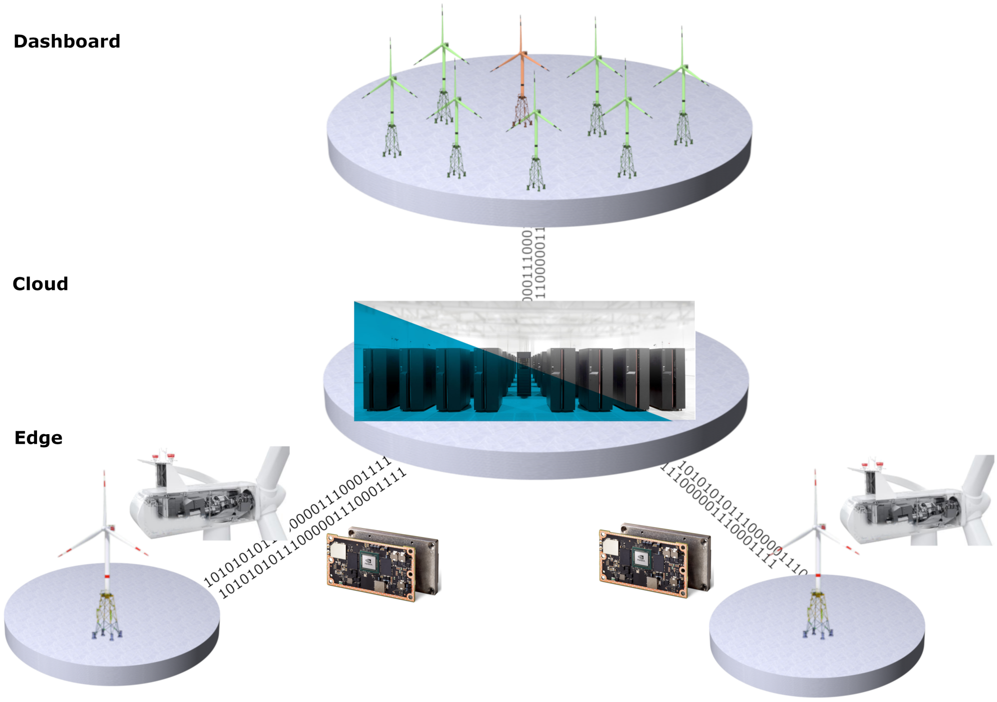


Figure 1 – Streaming data in the context of an edge-cloud balanced architecture for condition monitoring.

To overcome the streaming limitations, today, high frequency samples are typically taken for a short period of time in burst mode. Such bursts are done at intermittent periods in time. However, this means that for machines operating at non-constant speed and load chances are high that data-points are taken at different loading conditions. Due to the continuously changing nature of the system excitation, its response will also permanently change. These changes can have an influence on the resulting monitoring feature values. If the data samples—acquired at the intermittent moments—are spread too much over different loading conditions, then trending becomes challenging. Particularly since today's innovative industrial machines typically target a wider operational range, their speed and loads are varying continuously. To allow extraction of high quality condition indicators it becomes therefore interesting to explore, not only continuous data collection, but also continuous processing. This paper targets the latter.

Ideally we can instrument all machines in the fleet. The collected data should then allow the extraction of directly actionable insights for machine designers on the one hand machine and for owner-operators on the other hand. The one will use the insights to improve the design, whereas the other uses alarms to perform predictive maintenance. To allow the instrumentation of many machines it is necessary to have integrated processing algorithms capable of automatically processing the monitoring data. Edge computing can play an important rule to allow the extraction of health and design information from a large number of machines in a fleet where it might be unfeasible to transfer all data to a central location. Extensive research about the detection of failure in rotating machinery is available in literature today. More recently, machine learning is used more and more for condition monitoring. This offers opportunities towards automation. Leaning algorithms can enhance vibration signal processing methods to make them autonomous and more repetitive. This paper targets such methods by combining advanced signal processing techniques with anomaly detection and feature fusion based on data-driven techniques.

In this paper we target the assessment of the feasibility to use advanced edge devices for overcoming the limitations linked to intermittent data gap. We develop an integrated approach combining advanced signal processing methods with anomaly detection and a Bayesian regression approach to deal with vibration data in the new digital context. We target maximal computation close to where the sensor data is measured. Thus maximally leveraging processing power of the embedded ARM cores and GPUs. Devices of this architecture are plenty. In this paper, we use the NVIDIA Jetson TX2 embedded computing board, which features an ARM for general-purpose computations, and a GPU for specialized fast matrix-vector computations that are apparent in many machine learning techniques. This device combines low cost with high computational density. To assess the feasibility of using this device in a condition monitoring context, we use data of healthy and failed wind turbines from the multi-megawatt range.

2 Methodology

Our goal is to maximally automate the condition monitoring process such that calculations can be done maximally at the edge. Those approaches that cannot be calculated at the edge we will calculate centrally a the cloud level. As such we generate a balance between edge and cloud, as illustrated in Figure 1. For those approaches that cannot be calculated at the edge level, we target to only use high quality data.

We focus on the wind turbine drivetrain system. Different monitoring pipelines are calculated in parallel on the edge device. In this paper we assess the possibility to calculate processing pipelines of different complexity on the embedded processors. A multitude of features is generated. These range from statistical indicators on raw sensor data, that are less computationally intensive, to complex sequences of filters. Anomalies in these features are afterwards annotated using machine learning. To optimize the usage of the calculation power of the edge device, analysis methods of different types are coupled to their most optimal processor type. All signal processing calculations are done on the ARM processors. Multiple ARM processors in parallel allow to calculate features on multiple channels at the same time. The anomaly detection using Bayesian approaches is then done using the GPU processor. This allows to exploit the fast matrix-vector computations.

Since the focus of this paper is on the assessment of the edge computation aspect, we only discuss the pipelines used in this paper in a high-level overview. For details on the different methods the reader is for each sub-block referred to our prior publications or relevant literature. The following paragraphs discuss these processing pipeline blocks.

Statistical indicators
Statistical values of different nature can be calculated on acceleration data to detect changes in vibration behaviour of the system over time. We use the following indicators:

1. RMS: This gives an indication of the overall energy level present, $x_{RMS} = \sqrt{\frac{1}{N}\sum_n x^2(n)}$, with $x(n)$ the sampled signal.
2. Crest factor: Max peak value over RMS, $CF = \frac{|x_{peak}|}{x_{RMS}}$.
3. Kurtosis: A measure for the dispersion of the signal's distribution, $\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^2} - 3$.
4. Moors kurtosis: An alternative implementation of kurtosis based on quantiles [6], $\kappa_{Moors} = \frac{(E_7-E_5)+(E_3-E_1)}{E_6-E_2}$.
5. Peak-to-Peak: A straightforward indicator that quantifies the distance between the maximum and minimum acceleration, $x_{P2P} = x_{max} - x_{min}$.
6. Peak Energy Index: $PEI = \sqrt{\frac{1}{N_p}\sum_{n=1}^{N} x_p^2(n)}$, where N_p is the number of peaks exceeding a threshold equal to $\mu_x + 2\sigma_x$, with μ_x the mean and σ_x the standard deviation.

Speed compensation
Complex processing pipelines contain different processing steps that are chained together. For wind turbines a first step is always correction for speed fluctuations due to the stochastic nature of the wind. Typically this is achieved by converting the acceleration signals acquired in the time domain to the angular domain by means of angular re-sampling methods. Accurate speed measurements are necessary to achieve this step. Different methods are available in literature. We opt to use the Multi-Order Probabilistic Approach (MOPA). This method is based on interpreting the short time Fourier transform (STFT) of the vibration signal as a probability density function of the instantaneous angular speed. Consequently if the STFT has a high amplitude at frequency f,

then the probability that the shaft frequency is equal to f/H_i with H_i being the excitation order. For details on the specifics of the method the reader is referred to [4, 7].

Cepstrum liftering

In case bearing damage signature extraction is targeted, a second step consists in the removal of the dominating effect of the harmonics originating from the gears. Typically cepstrum liftering is used to achieve this goal. In this paper we opt to use the automated cepstrum editing procedure (ACEP). The cepstrum allows to concentrate the energy of periodically spaced spectral components into a smaller number of impulses. These are referred to as rahmonics. Removing these peaks decreases the amplitude of the periodic signal content. Similar to filtering in the frequency domain the term liftering is used in the cepstral domain. Randall & Sawalhi showed that the real cepstrum can be used to edit the log amplitude spectrum which contains the discrete harmonics [8]. Recombining this modified amplitude spectrum with the original phase creates an edited version of the time signal with significantly less pronounced harmonic content. In our case we use an automated cepstrum editing procedure. In this procedure first a long-pass lifter is applied to the cepstrum. The corresponding content will be read to the signal after the editing procedure. This in order to prevent liftering of this content which is dominated by the structural behaviour of the system. First noise reduction is achieved by a wavelet de-noising approach. Then a comb lifter allows the removal of selected distinct peaks in the cepstrum. Finally the signal is transferred back to the time domain. In the resulting signal the stochastic content is dominant.

Filtering

Once the data is cleaned and disturbances are removed, filtering is done prior to calculation of the statistical indicators. A variety of frequency ranges and filter types can be chosen. For the purpose of this paper the frequency range up to the Nyquist frequency is divided in 4 different bands. For each of these bands the same statistical indicators as for the raw data are calculated.

Cyclic spectral coherence

In parallel to the pipelines calculating statistical features, enveloping is done to detect bearing faults. Even after reducing the influence of harmonic disturbances using cepstrum liftering techniques the detection of bearing faults remains a challenge. The bearing fault impulsive excitation signatures are small compared to potential other disturbing energy sources. In complex gearboxes with many stages this can complicate detection. Enveloping techniques linked to band pass filtering allow to highlight the fault. However, accurate knowledge about the most optimal frequency band for envelope extraction is needed. The use of cyclic spectral coherence to identify this frequency band improves detection potential by identifying the frequency ranges in which resonances can amplify the signatures [2]. Calculation of the spectral coherence maps and envelopes are done both directly at the edge.

Anomaly detection

The features that result from the processing pipelines discussed above are treated as time series data. This allows them to be trended in order to accurately capture failure-driven anomalies. To automate and objectify this process, we use linear Bayesian Ridge regression [5] to model the features under healthy conditions using the machine operational parameters as inputs. Bayesian Ridge Regression is a probabilistic approach to regression with regularization. Essentially, it fits the linear parameters (i.e., slopes and intercept) and inherent noise to the observed data, while maintaining the uncertainty over the parameters. This type of regression fully captures all stochastic components in the modeling process, and allows the identification of anomalies that are not due to this stochasticity. The models are trained during a healthy period and thus predict expected feature response for healthy conditions. Outliers exceeding a three standard deviations range around the expected feature behaviour are classified as anomalies and thus unhealthy behavior.

3 Experimental case

To be able to keep up with the continuous data streams we opt to not transmit all data to a central cloud processing platform but balance processing between cloud and edge. Figure 1 gives a schematic overview of our architecture. In this paper we will make use of embedded ARM and GPU technology and use the commercially available NVIDIA Jetson TX2 product as edge device. This features a quad-core ARM Cortex-57 MP Core CPU. In addition also a 256-Core NVIDIA Pascal GPU is available.

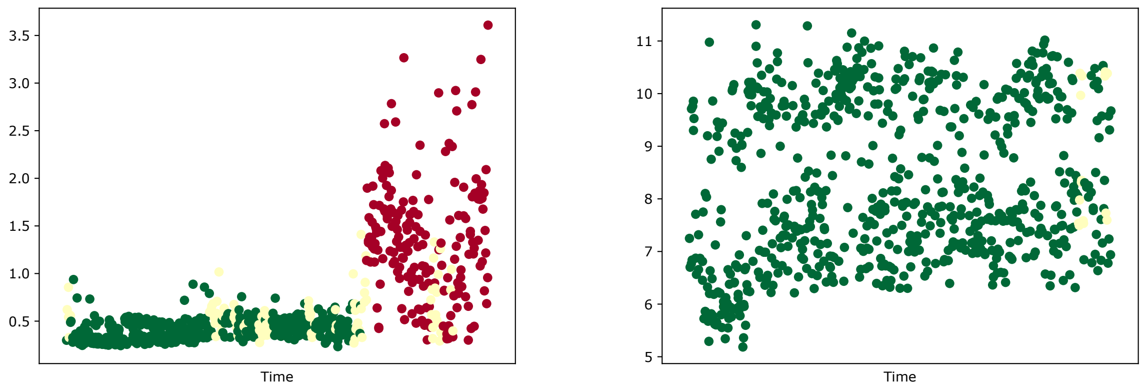
Signal processing pipelines exploit a Python architecture. These are deployed on the quad-core ARM processor. The anomaly detection models are implemented in TensorFlow [1], which allows for execution at the edge on the GPU. TensorFlow is a library that automatically organizes operations in a computational graph, such that many repetitive simple operations can be executed in parallel on the arithmetic units of the GPU. Such a framework is suitable, as Bayesian Ridge regression requires matrix multiplications during prediction, and thus can be efficiently executed on a GPU. Training of the models is performed in the cloud, as it is too computationally intensive.

To asses the processing ability of the Jetson and identify its limitations we opt to use vibration monitoring data collected from offshore wind turbines and process the analysis pipelines discussed above. Based on the insights gained in the speed and processing capabilities of the ARM processor and GPU, the balance between edge and cloud computing is identified. Computations that are feasible to be performed at the edge are done there, whereas the other processing is done in the cloud. To limit bandwidth usage of the data connections only high quality data is sent to the central cloud. The indicators calculated at the edge allow to determine whether it is useful to transfer the data to the central cloud.

Industrial CMS data is taken as starting point. In this way the analysis is done on a representative dataset. Data-sampling rates are above 25kHz. Each data block is approximately 10 seconds. Data is collected at intermittent moments in time. One accelerometer channel at the planetary and one at the high speed stage are processed. First, the processing pipelines using signal processing techniques, discussed in the previous section, are calculated. The six statistical indicators are calculated on the raw sensor data as a baseline. 240 processing pipelines provide an extensive feature set. The final step in each pipeline is the calculation of a statistical indicator to generate a summary value. To allow data-trending, intermittent data samples over a multi-year period are processed. After the trending step anomaly detection is executed on the GPU of the Jetson for fast prediction. Model training, comparison, and visualization of the resulting features is done on the VUB AVRG cloud platform.

Accelerometers are mounted on the gearbox at the planetary and high speed stage. We target the detection of deterioration on the planetary gear stage, which can lead to more severe damage. Constructed health indicators of both the planetary stage and high-speed stage are compared to illustrate deviations in nominal behavior. Figure 2 shows an example, comparing a peak-to-peak feature for the planetary stage and the corresponding feature for the high speed stage over multiple years of data. Based on these indicators it is clear that the fault is in the planetary stage. The indicators clearly show a strong anomaly score towards the end, whereas before some outliers already start to pop up. The indicators for the high speed stage show no anomalous behaviour. This shows the ability to locate the fault in the gearbox system.

Based on this analysis it is possible to perform calculations for failure detection, taking into account a large quantity of indicators calculated in parallel at the edge. For the moment calculations are far from real-time so only intermittent measurements are possible. As such more computationally intensive calculations using more detailed processing methods, such as for example the Kurtogram [3], need to be performed at the cloud level anyhow. However, there is definitely potential to use this technology for continuous condition monitoring if enough calculation cores are made available on the device.



(a) Planetary stage

(b) High-speed stage

Figure 2 – Comparison of peak to peak based health indicator between two affected and unaffected channels.

4 Conclusion

This paper investigated the potential to use combined embedded ARM and GPU processing architectures for edge computing in the context of condition monitoring. Both signal processing and machine learning approaches were calculate locally on the device. The NVIDIA Jetson TX2 was used as testing device. Using real-life data it was shown that failure detection can be achieved by edge computing. Complex signal processing pipelines, comprising of amongst others speed compensation, cepstrum liftering and enhanced enveloping were calculated on the device. These was complemented with Bayersian feature fusion using Tensor Flow on the embedded GPU after model training in the cloud.

As such it is shown that the edge device can be used to monitor a gearbox using typical measurements from CMS devices used in industry today. In addition to these local calculations the computationally more intensive calculations able to detect this failure type earlier will be performed in the cloud. In future research this balance will be further optimized.

Acknowledgements

The authors would like to acknowledge FWO (Fonds Weten- schappelijk Onderzoek - Vlaanderen) for their support through the SB grant of Timothy Verstraeten (1S47617N), the agency for Innovation by Science and Technology in Belgium for supporting the SBO HYMOP project

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] J. Antoni. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and vibration*, 304(3-5):497–529, 2007.

[3] J. Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007.

[4] Q. Leclere, H. André, and J. Antoni. A multi-order probabilistic approach for instantaneous angular speed tracking debriefing of the cmmno 14 diagnosis contest. *Mechanical Systems and Signal Processing*, 81:375–386, 2016.

[5] D. J. C. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992.

[6] J. Moors. A quantile alternative for kurtosis. *The statistician*, pages 25–32, 1988.

[7] C. Peeters, Q. Leclere, J. Antoni, P. Guillaume, and J. Helsen. Vibration-based angular speed estimation for multi-stage wind turbine gearboxes. *Journal of Physics: Conference Series*, 842(1):012053, 2017.

[8] R. B. Randall and N. Sawalhi. A new method for separating discrete components from a signal. *Sound and Vibration*, 45(5):6, 2011.