ANALYTICAL ALGORITHMS

Michael Barth and Michael Kasch, ILF, Germany, explain how artificial intelligence can help operators find hidden correlations and patterns in their data.

Buzzwords like 'data mining', 'deep learning' and 'artificial intelligence' (AI) are currently flying around in almost every business and industry area. The same technologies used for screening users' privacy and analysing customers' preferences to improve sales figures are also applied to optimise technical processes for maintenance of assets and, with constantly growing attention, for physical and cyber protection of critical infrastructure. This article presents an overview of the methods currently being used, and outlines opportunities to add value to the safe and efficient operation of pipelines by applying AI methods to their data, which are rapidly growing as pipeline systems and facilities are operated, monitored and controlled in 24/7 mode.

Pipelines are equipped with sensors, actuators, specific automation systems and control loops. Sensor data, being continuously interrogated with the aim of monitoring process

equipment, function units and systems of pipelines, are typically collected in real-time, and these data are transmitted via dedicated process networks to control systems with accordingly high frequency. It is not unusual for a control centre to receive hundreds or even thousands of real-time data each second from remotely controlled stations. Many of these data are being archived for different reasons, from purely a technical purpose to commercial reporting to legal requirements. But compared to the huge amount of data which is inevitably accumulating over time, while operating 24/7, almost nothing of this potential source of information is actually used to benefit the operating organisation. The full picture of process conditions does not only include measured process data but also the control signals being sent in the opposite direction to controllers and actuators – either manually by human operators or automatically by automation systems. Process controllers are continuously gathering process information, evaluating the present process states and ultimately sending control signals to steer process parameters towards their set-points, and in this way keeping the systems within their designed operating envelopes.

All these data bear valuable information that can be analysed by AI algorithms. Unknown correlations and hidden dependencies between process and status parameters can be revealed and then used for process optimisation and the reduction of risks.

External databases may contribute valuable additional information, and can be included in the analysis to spot correlations between internal and external data such as weather or economy statistics, or geographical and topographical information. Both business applications and process control systems are being attacked, and data sources might be compromised. Thus, internal data and systems are to be protected by the defence-in-depth approach to prevent manipulation and disclosure.

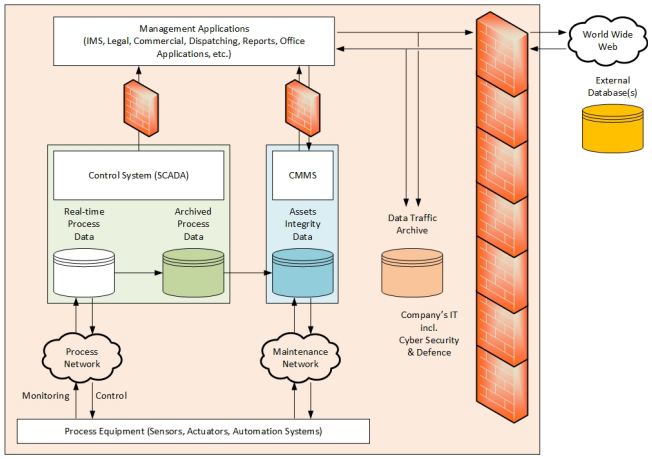

Figure 1. Typical architecture for monitoring and control systems for process assets.

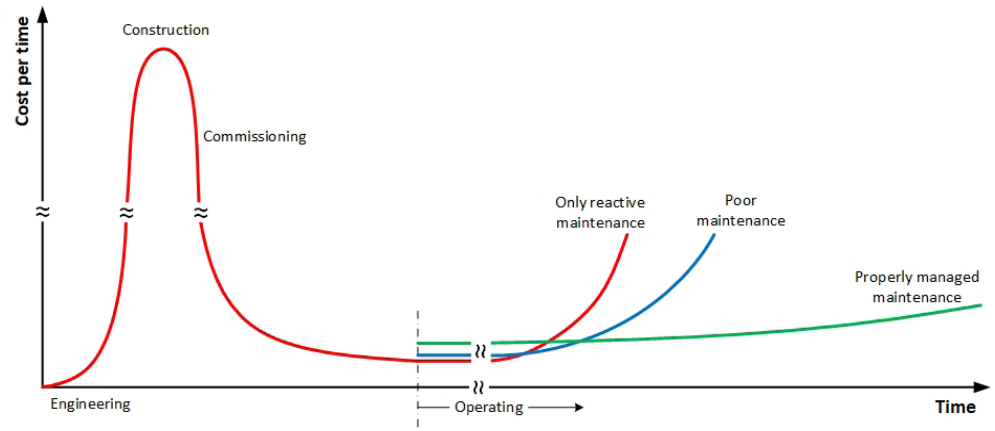

Figure 2. Cost control through proper maintenance management.

Methods

There are two distinct approaches to analyse big data volumes. The top-down approach is driven by a hypothesis. The data pool is combed through with a defined expectation of what patterns or signatures might be found therein.

The other method is the opposite, called the bottom-up approach. It is data-driven, and crawling through the data jungle is performed without a hypothesis or clear expectation of what could be found. The latter approach is used to reveal hidden patterns and unknown correlations between data of different kind and origin. If something is found by bottom-up data analysis, it will at least be a surprise. It is then the challenge to use this new, unexpected information or regularity to improve process efficiency, safety and security, and to reduce potential risks.

Three main methods for data mining are used: association, classification and clustering. Irrespective of the method used, raw data needs to be preprocessed before data mining methods can be efficiently applied to big data volumes. This is particularly necessary if different data sources are to be included. The effort of preprocessing data should not be underestimated. It normally takes more than 50% of the entire exercise.

- Association methods uncover hidden dependencies between objects which are represented in databases in terms of their attributes and variables. Patterns uncovered by association can typically be formulated as IF-THEN relations.
- Classification methods are looking for concordance or at least similarity of data and data sets with predefined and given target patterns.
- Clustering is a type of classification method which can autonomously identify patterns (if existing) without any predefined target. Clustering algorithms can reveal hidden and unknown relationships and dependencies in data pools.

Apart from data mining methods, there is a continuously

growing number of AI algorithms. Deep learning is a class of optimisation methods for artificial neural networks with multiple 'hidden' layers of neurons between the input and output (result) interface layers. Such neural networks are characterised by a high level of internal complexity and are called 'deep' in terms of their large number of hidden layers. Such methods are typically used for machine learning, pattern and face recognition, real-time language processing and other complex challenges. Deep learning methods are based on self-adapting algorithms, and they need a learning or training phase to become effective and reliable.

Potential applications

The following sections present a small selection of possible applications for data mining and deep learning in the process industry, but can also be used for inspiration in identifying organisation-specific topics and potential tasks that employ AI technology.

Pipeline corridor monitoring

A technology called distributed acoustic sensing (DAS) allows an operator to listen in to a fibre optic cable (FOC) over long distances. Laser pulses are sent through one of the fibres inside the FOC. The tiny portion of backscattered light carries information about the mechanical stress/strain conditions of the fibre.[1] Bending the FOC or vibrations imposed on it cause disturbance on the local backscattering characteristics of the fibre. Such anomalies can be detected and precisely located by using optical time domain reflectometer (OTDR) technology. The fibre functions as a spatially distributed sensor for static and dynamic mechanical stress, and can therefore be regarded as a spatially distributed microphone over distances of currently up to 50 km with a single DAS interrogator unit. It is worth noting that DAS systems use standard fibres of ordinary telecommunication FOCs, and no special sensor fibres are required.

Fibre optic cables are typically installed along pipelines to establish communication between pipeline stations and control centres. When installed close to the pipeline, preferably in the same trench, DAS can continuously monitor the pipeline corridor for acoustic anomalies. Earthworks, manual digging, landslides and earthquakes cause acoustic signatures in the ground which can be detected and localised by DAS.

Software algorithms are used to evaluate the complex signal patterns received by the DAS interrogator. AI algorithms can substantially improve the evaluation process by deep learning and finally by autonomously interpreting the backscattered light patterns for alarming, in case of third-party activities on the pipeline corridor.

When properly designed and installed, DAS systems can even detect and precisely locate leaks due to their characteristic sound patterns in high pressure pipelines.

Perimeter security

DAS technology can also be used to monitor and secure the perimeter of pipeline stations. The FOC can either be buried or mounted to the fence and used for physical intrusion detection. When installed underground, the FOC is the DAS sensor for detecting vehicles or persons approaching or trespassing the secured area. Climbing or cutting the fence, as well as digging underneath, is detected and located by the same method.

Cyber security

Cyberattacks on industrial control systems or critical infrastructure bear enormous social and economic risks. Many of the traditional IT security approaches and techniques used today require prior knowledge of the patterns or signatures of potentially compromising data packages in the incoming data stream (classification). In many cases, these tools (or the way they were deployed) failed to successfully defend systems from cyber intrusion and clandestine installation of malware.

With the constantly growing computing power and improving AI algorithms, new tools are emerging which are capable of analysing network traffic almost in real-time. Such advanced systems use AI algorithms which autonomously learn to identify potentially compromising patterns in the data stream (clustering). This capability is an indispensable precondition for fast and effective responses to cyberattacks.

Process and integrity monitoring

Pipelines are monitored constantly. Core process data are typically visualised at control centres in order to provide the operators there with comprehensive, but easy to grasp, real-time status information. Normal operating states yield characteristic patterns of process data. Deviations from normal operation will result in deviating patterns that can be spotted by algorithms much faster than by human operators, especially for process data which are not permanently visualised. Such algorithms need to learn how abnormal process states appear, and they are rare under normal circumstances. Therefore, it may take weeks, months or even longer before AI can reliably support the control room operators.

Process simulation could be a valuable measure to shorten the learning phase.[2] Feeding the learning algorithm with the simulated process data of operating scenarios which do not normally occur or would not be intentionally provoked, even for testing purposes, would reduce the learning phase.

A prudently trained AI algorithm can be regarded as an equivalent to a driving assistant on a car. The support for the operator can range from hints, reminders and messages up to serious alarms. Conventional hydraulic-based leak detection systems (LDS) are a good example for such an improvement potential.

A well known flaw of all hydraulic-based LDS is their sensitivity. The limitations of such systems, which cannot be circumvented, are given through the accuracy of instrumentation, the underlying hydraulic modelling as well as system-specific factors.

Many leaks with severe consequences remain undetected for a long period of time, even if equipped with the best available instruments. No alarm is raised in due time, and this makes many pipeline operators conclude that LDS are of very limited use. But is this the right conclusion? It rather reflects the gap between expectation (driven by overly optimistic promises)

and the principle capabilities of conventional hydraulic or SCADA-based LDS.

Imagine a perfect LDS which can technically detect leaks at 0.25% of design throughput. To avoid false alarms, the alarm threshold selected may be 0.5%. Perfect LDS then means that a leak of 0.49% of design throughput would not raise an alarm, and this in full compliance with its specification. In other words, the LDS works as per approved specification. However, the patterns of process data indicating leakage would be very similar to a situation with a slightly higher leak rate, for which an alarm would be raised. AI technology can substantially reduce the safety margin between technical sensitivity and the practically chosen alarm threshold.

Computerised maintenance management systems

Currently, maintenance of assets and the integrity management for assets is risk based and includes proactive condition monitoring. All asset data are administered with computerised maintenance management systems (CMMS). The CMMS is a software tool based on a database hosting all available data and condition information about each piece of equipment, which is continuously updated according to its utilisation and updated after maintenance or repair activities; hence, it is a big data volume which is constantly growing. External statistics on the utilised equipment and instrumentation can be included in condition evaluation processes.

As every piece of equipment is interacting with other equipment, there are quite obvious interdependencies. However, there might be hidden dependencies which can be uncovered by clustering and association.

Maintenance management, as part of integrity and lifecycle management, is essential for cost control, safe and efficient operations and ensuring operational availability. Typical cost developments over the lifetime of pipeline systems with respect to different maintenance approaches is illustrated in Figure 2.

Unknown optimisation potential for maintenance can be uncovered using AI methods, and in this way can contribute to both mid and long-term cost reduction.

Summary

Steady advances in AI and constantly growing computing power are advancing into almost every realm today. Many more potential uses than those outlined in this article can and will be identified for AI applications. Wherever analysis and evaluation of complex situations or patterns is required, AI can provide valuable support. Uses for pipeline operators extend from control room operating (SCADA), integrity monitoring and maintenance, to physical and cyber security. WP

Notes

1. Analogue to the presumably better known fibre optic distributed temperature sensing (DTS).
2. Offline simulation without any feedback into the productive process systems.