Active anomaly detection and prediction in Industry 4.0

Francesco Amigoni[1], Nicola Basilico[2], Alessandro Farinelli[3], Luca Iocchi[4], Andrea Lanzi[2]
[1]Politecnico di Milano, [2]Università degli Studi di Milano,
[3]Università degli Studi di Verona, [4]Sapienza Università di Roma
francesco.amigoni@polimi.it, nicola.basilico@unimi.it, alessandro.farinelli@univr.it,
iocchi@diag.uniroma1.it, andrea.lanzi@unimi.it

Abstract

This report introduces a joint research effort of 4 AIIS research labs focused on active anomaly detection in industrial scenarios, thus targeting the *AI for Industry* topic. The main goals and challenges of active anomaly detection are briefly described, as well as the expected results and impact of this research line. Some preliminary work done by the research labs is also briefly discussed.

1 Introduction

The increasing use of digital technologies is rapidly changing the production systems worldwide, resulting in a fourth industrial revolution, dubbed Industry 4.0. Artificial Intelligence (AI) is considered a key driver for Industry 4.0 as it can increase the quality of production, reducing costs and improving the condition of human labour.

This research line focuses on *anomaly detection*, a key topic for Industry 4.0, and builds on the recent scientific and practical trend that focuses on data-driven methods and advanced machine learning techniques to detect anomalies. These methods aim at building a model by observing the data generated by the system operating in nominal conditions. The model is then used to determine whether the current system state exhibits behaviours that do not adhere to an expected/nominal pattern (i.e., anomalies). These methods are mostly passive observers of the system and usually concentrate on current anomalies without propagating the analysis to possible future states.

Our aim is to go beyond the current concept of anomaly detection to address an increased level of "intelligence" (4.x) for Industry 4.0. In particular: i) we will investigate *active* AI algorithms that plan and perform explicit actions (e.g., move a camera or a robot carrying sensors) to improve the accuracy of anomaly detection, to obtain better information with fewer resources, and to enable collaboration with human operators (e.g., by suggesting and explaining useful actions to be performed); ii) we will develop anomaly *prediction* algorithms that can identify vulnerabilities in estimated future states of the system. This is a very important trend for Industry 4.0, where predictive technologies for system maintenance is a key topic.

We will consider anomalies as consequences of two different root causes: 1) machine failures and 2) adversarial cyber-attacks strategically engineered to jeopardize the industrial process without being detected. The latter represents a subtle and compelling form of anomaly for complex industrial systems that is attracting increasing interest from the cyber-security community. The active anomaly detection and prediction algorithms will be based on a combination of machine learning and symbolic decision making.

2 Related work

This section overviews the scientific state of the art on AI methods for anomaly detection in industrial systems and discusses some industrial practices.

2.1 AI methods for anomaly detection

Following the literature [Aoudi *et al.*, 2018], by "anomaly" we mean a deviation from the expected behavior of a system, independently of the reason (fault or attack). The techniques we consider will address malicious attacks and unexpected events that may hinder the correct execution of standard production processes. We will focus on attacks or events that affect the particular nature of complex industrial systems, notably their interactions with the physical environment.

Anomaly detection approaches for complex systems that interact with the real world can be divided into three broad categories: model-based, knowledge-based, and data-driven [Chandola *et al.*, 2009; Khalastchi e Kalech, 2018]. Model-based approaches require explicit analytical models (i.e., mathematical equations) of systems and therefore need expert knowledge to be built. Knowledge-based approaches typically associate each known fault with a detection rule triggered when a specific behavior is observed. Data-driven approaches are based on (usually probabilistic) descriptions of behaviors or faults that are automatically learned from the system's observations. Their advantage is that they do not need any explicit prior knowledge of the system and of the faults.

Our focus is mainly on data-driven approaches. Specifically, online data-driven methods are typically used for complex industrial systems that often operate on time series (e.g., stream of sensor data), generating a probabilistic model of systems' behaviors in real-time. This model is then used to distinguish potential anomalies from nominal behaviors.

Some approaches (e.g., [Christensen *et al.*, 2008]) adopt supervised machine learning methods to classify data acquired in real-time. However, supervised methods need fully labeled data for training, which are not always available. Hence, recent developments shift to unsupervised and semi-supervised learning. Unsupervised methods (e.g., [Khalastchi *et al.*, 2015]) do not require a labeled training set but rely on the assumption that anomalies are rarely occurring. Semi-supervised methods (e.g., [Park *et al.*, 2016; Park *et al.*, 2017; Park *et al.*, 2019]) relax this assumption, but require labeled instances for the nominal class (that are usually easier to collect with respect to anomalous ones) [Chandola *et al.*, 2009].

Deep learning models have recently been employed to readdress several spatio-temporal modeling tasks, like those relative to anomaly detection, providing significant improvements over classical methods. The most used models include Autoencoders (AEs) [Hinton e Salakhutdinov, 2006] and Variational Autoencoders (VAEs) [Kingma e Welling, 2014]. AEs are particular kinds of artificial neural networks, which are trained to represent their input in a latent space and to reconstruct it. VAEs are similar but assume the existence of a probabilistic model parameterized by a latent variable that generates the observed input values. AEs have been widely used for anomaly detection on time series coming from systems that interact with the real world [Malhotra *et al.*, 2016; Zhang *et al.*, 2018]. More recent proposals include an LSTM-VAE using a reconstruction-based anomaly score and a state-based threshold to detect anomalies for robot-assisted feeding [Park *et al.*, 2017] and applying the STORN model [Bayer e Osendorfer, 2014] to anomaly detection by introducing a trending prior on the latent representation [Soelch *et al.*, 2016]. In [Chen *et al.*, 2020], a sliding-window VAE is proposed, which performs real-time anomaly detection on multivariate time series acquired from an industrial robot. Another example employs a VAE with attention [Pereira e Silveirat, 2018].

All these approaches build a model from the system's data and apply it to the current system state to detect anomalies. In this sense, these approaches are passive observers of the system and concentrate on current anomalies without propagating the analysis to possible future states. Against this scenario, we will develop active AI methodologies that can undertake actions to improve the accuracy of anomaly detection requiring fewer observations and possibly collaborating with human operators. We will also focus on anomaly prediction by devising methods for identifying vulnerabilities in estimated future states of the system. This challenge is in line with the recent trends in Industry 4.0 that recognize the positive impact of predictive technology for system maintenance [Selcuk, 2017].

2.2 Industrial practices for anomaly detection

ICT solutions currently employed in industries vary significantly based on market sector and size, but are typically based on a standard automation pyramid organized in 5 main levels: equipment (e.g., the devices), control (PLC, HMI), supervision (SCADA), operation (Manufacturing Execution System, MES) and management (Enterprise Resource Planning, ERP). Anomaly detection is typically addressed by the Supervisory Control and Data Acquisition (SCADA) system. SCADA systems usually focus on the operation of specific components (e.g., a 3D printer or an automated storage system) reporting possible anomalies for the operation of the single component (e.g., a user is interacting with the storage system and hence the system should not move to avoid hurting the user).

The development of anomaly detection and prediction systems that acquire data from the different SCADA units and operate at a higher level is recognized as an important and interesting direction but there are few market-ready products operating at this level. For example, Siemens proposes the use of Mindsphere, a cloud-based platform that collects information from the field through specific components (Mindconnect) and can perform data analytics. On the same line, the Industrial Anomaly Detection component aims at detecting anomalies in communication activities to avoid standard attacks (such as man-in-the-middle or denial-of-service).

While AI techniques (in particular, machine learning) are well known in the scientific community to be able to solve anomaly detection and prediction, these techniques are not yet available as off-the-shelf solutions in Industry 4.0.

We thus aim at bringing two important innovations based on AI techniques in Industry 4.0 anomaly detection and prediction: 1) the development of semi-supervised and unsupervised machine learning models for detection and prediction of anomalies, possibly coming from cyber-attacks; 2) the development of active strategies to better detect the anomalies and possibly to mitigate or resolve them.

To further stimulate the evolution of advanced manufacturing solutions, there have been initiatives that mention the term Industry 4.1 [Cheng *et al.*, 2016] and Industry 5.0[1]. In this perspective, we use the term Industry 4.x to indicate the investigation of new solutions that go beyond the current standards of Industry 4.0 but that are applicable to ICT solutions currently in place in the production systems.

3 Expected advancements

Data-driven anomaly detection represents a well-established approach for complex systems security and reliability whose high potential and effectiveness are driving significant efforts from both the cyber-security and AI research communities. The fast-paced advancements of Industry 4.0 provide a ground where such technologies could thrive. Increased automation, enabled by smart and pervasive sensing and communication technologies, comes with a richness of real-time data that is already fostering intelligence and autonomy at large, with particular emphasis on predictive technologies. Autonomous or assisted production systems, process optimization, synthesis of digital twins, and predictive maintenance are examples of technological setups where AI can play a prominent role and where guaranteeing expected/nominal functioning is critical for the whole industrial process. Large and heterogeneous amounts of data represent an opportunity for anomaly detection and prediction, intended as the process

[1] https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/industry-50_en

of identifying/predicting behaviors that do not belong to an expected/nominal pattern.

As already mentioned, we consider anomalies in an Industry 4.0 environment as consequences of two distinguished root causes: 1) early signs of machine failure or equipment breakdown, 2) adversarial cyber-attacks strategically engineered to jeopardize the industrial process without being detected.

In contrast with anomaly detection approaches currently in use by industries that focus on monitoring specific components and are tailored to identify known issues and faults that are most likely to occur, we expect AI-based approaches to monitor the whole production system.

More specifically, we envision a high-impact research line on novel AI-driven methods for active anomaly detection and prediction in Industry 4.0. A first key contribution will focus on the *activeness* of the proposed methods. Being active entails the capability of making decisions on a state representation of the system. Such a state is built by extending a model of the system's high-level dynamics with on-line knowledge persistently gathered and extracted from data. Leveraging such a representation, an active anomaly detection system proactively makes decisions about where to acquire data (performing data selection or reconfiguring the data collection subsystems), how to adapt detection/prediction thresholds, and what actions to undertake in the physical environment to stop, mitigate, or prevent an identified threat to the industrial process. A second key contribution will involve the step from anomaly detection to anomaly *prediction*. An anomaly prediction system is capable of forward propagation of the system's current state, by estimating possible future transitions in accordance with the data and the system's dynamics. Anomaly detection can then be applied to a set of probable future states to identify vulnerabilities that can translate in future attack opportunities.

By introducing the active and predictive dimensions, we aim at achieving an increased level of AI-based tools for anomaly detection and prediction in Industry 4.0 systems. We expect that these tools will properly combine cutting-edge machine learning methods with knowledge representation and decision making in a unified framework. Although we will instantiate the problem of anomaly detection and prediction to Industry 4.0 applications, it can be of interest for many other applications, including for example all those related to security and protection of critical infrastructures.

While anomalies due to malfunctions of production systems are relatively well studied, attacks represent a more subtle and interesting form of anomaly to detect and predict in complex industrial systems and this research will also focus on them. Our aim is to construct threat models in the scope of two reference scenarios. The first one involves Pick-up and Delivery (PD) tasks with any level of automation. The second one pertains to automated Quality Control (QC) performed with vision-based systems. These scenarios include basic, widely deployed, and highly interconnected sub-processes for a great variety of Industry 4.0 manufacturing plants and represent fundamental building blocks of the industrial process where damages or breakdowns can have remarkable costs. In PD, adversarial attacks can undermine efficiency by slowing down the transportation of components or by inducing congestion altering the workload demand in time. In QC, attacks can degrade the effectiveness of quality control by feeding the vision-based system malformed inputs obtained via environmental disturbances or even injecting alterations of the product itself. In both scenarios, an active system can introduce remarkable benefits.

We will deal with the difficulty in obtaining training data for active systems for which static datasets are not enough. So, a significant advancement with respect to the state of the art will be the development of techniques that combine datasets, simulations, and real-world experiments to collect and augment data for training active anomaly detection and prediction methods. Then, we will pursue a leap from AI algorithms that perform mostly passive anomaly detection to AI algorithms that are active and that can also perform anomaly prediction. The proposed solutions will be extensively tested both in simulated and real industrial systems, demonstrating the benefits they can provide.

Overall, the above mentioned advancements will provide a paradigm shift for an important research topic that is characterized, more in general, by the integration of symbolic AI-based techniques with data-driven machine learning approaches. Moreover, the deployment of active anomaly detection approaches in representative examples of Industry 4.0 scenarios will show how the scientific research can be immediately and directly applicable to important industrial use cases, thus having a strong and deep impact also on the industrial community.

4 Some preliminary results

In what follows we briefly describe some of our recent works that we consider important stepping stones towards the goals we outlined above.

We have considered the goal of monitoring the behaviors of robotic platforms to identify unexpected or anomalous behaviors. The main idea is to address the problem of anomaly detection by analysing data and specifically, time series, acquired by sensors and actuators during the routine operation of the robotic platforms.

In [Olivato *et al.*, 2019], an approach that converts sensor logs into images and then uses a convolutional AE to detect anomalies caused by cyber-security attacks is proposed. The method has been extended in [Brigato *et al.*, 2021] comparing different representations of sensors logs and different neural network architectures. The results show the effectiveness of anomaly detection based on sensor logs time series.

In [Azzalini *et al.*, 2020], we model the nominal (expected) behavior of a robot system using Hidden Markov Models (HMMs) and we evaluate how far the observed behavior is from the nominal one using variants of the Hellinger distance. The use of the Hellinger distance positively impacts on both detection performance and interpretability of detected anomalies, as shown by results of experiments performed in two real-world application domains, namely, water monitoring with aquatic drones and socially assistive robots for elders living at home.

In [Azzalini *et al.*, 2021], we model the behavior of the robot systems using a VAE architecture able to model very long multivariate sensor logs exploiting a new incremental training method, which induces a progress-based latent space that can be used to detect anomalies both at runtime and offline. The VAE is trained using unlabeled observations of a robot performing a task, containing both nominal and anomalous executions. Only a very little amount (even just one) of labeled nominal executions is then required to partition the learned latent space into nominal and anomalous regions. Experimental results show that our method outperforms state-of-the-art anomaly detectors commonly used in robotics both in terms of false positive rate and alert delay.

While the results of the above mentioned works are certainly very promising and have confirmed the effectiveness of the used techniques, additional research is indeed required in order to fully achieve the ambitious objectives described in this paper.

Among others, we just mention here an important open challenge that is robustness against adversarial attacks. An unpublished work on this topic confirms that several anomaly detection components can be attacked by generating malicious behaviors that are not detected as anomalies.

We believe that an effective integration of several AI techniques, properly combining data-driven and model-based reasoning components, can produce effective and robust active anomaly detection and prediction methods for Industry 4.0.

References

[Aoudi *et al.*, 2018] W. Aoudi, M. Iturbe, e M. Almgren. Truth will out: Departure-based process-level detection of stealthy attacks on control systems. In *Proc. CCS*, 2018.

[Azzalini *et al.*, 2020] D. Azzalini, A. Castellini, M. Luperto, A. Farinelli, e F. Amigoni. HMMs for anomaly detection in autonomous robots. In *Proc. AAMAS*, 2020.

[Azzalini *et al.*, 2021] Davide Azzalini, Luca Bonali, e Francesco Amigoni. A minimally supervised approach based on variational autoencoders for anomaly detection in autonomous robots. *IEEE RA-L*, 6(2), 2021.

[Bayer e Osendorfer, 2014] J. Bayer e C. Osendorfer. Learning stochastic recurrent networks. arxiv.org/abs/1411.7610, 2014.

[Brigato *et al.*, 2021] Lorenzo Brigato, Riccardo Sartea, Stefano Simonazzi, Alessandro Farinelli, Luca Iocchi, e Christian Napoli. Exploiting time dynamics for one-class and open-set anomaly detection. In *Artificial Intelligence and Soft Computing - 20th International Conference, ICAISC*, volume 12855 of *LNCS*, pages 137–148. Springer, 2021.

[Chandola *et al.*, 2009] V. Chandola, A. Banerjee, e V. Kumar. Anomaly detection: A survey. *ACM Comput Surv*, 41(3), 2009.

[Chen *et al.*, 2020] T. Chen, X. Liu, B. Xia, W. Wang, e Y. Lai. Unsupervised anomaly detection of industrial robots using sliding-window convolutional variational autoencoder. *IEEE Access*, 8, 2020.

[Cheng *et al.*, 2016] F. Cheng *et al.* Industry 4.1 for wheel machining automation. *IEEE RA-L*, 1(1), 2016.

[Christensen *et al.*, 2008] A. Christensen, R. O'Grady, M. Birattari, e M. Dorigo. Fault detection in autonomous robots based on fault injection and learning. *Auton Robot*, 24(1), 2008.

[Hinton e Salakhutdinov, 2006] G. Hinton e R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 2006.

[Khalastchi *et al.*, 2015] E. Khalastchi, M. Kalech, G. Kaminka, e R. Lin. Online data-driven anomaly detection in autonomous robots. *Knowl Inf Syst*, 43(3), 2015.

[Khalastchi e Kalech, 2018] E. Khalastchi e M. Kalech. On fault detection and diagnosis in robotic systems. *ACM Comput Surv*, 51(1), 2018.

[Kingma e Welling, 2014] D. Kingma e M. Welling. Auto-encoding variational Bayes. In *Proc. ICLR*, 2014.

[Malhotra *et al.*, 2016] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, e G. Shroff. LSTM-based encoder-decoder for multi-sensor anomaly detection. In *Proc. ICML Anomaly Detection Workshop*, 2016.

[Olivato *et al.*, 2019] M. Olivato, O. Cotugno, L. Brigato, D. Bloisi, A. Farinelli, e L. Iocchi. A comparative analysis on the use of autoencoders for robot security anomaly detection. In *Proc. IROS*, 2019.

[Park *et al.*, 2016] D. Park, Z. Erickson, T. Bhattacharjee, e C. Kemp. Multimodal execution monitoring for anomaly detection during robot manipulation. In *Proc. ICRA*, 2016.

[Park *et al.*, 2017] D. Park, Y. Hoshi, e C. Kemp. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. *IEEE RA-L*, 3(3), 2017.

[Park *et al.*, 2019] D. Park, H. Kim, e C. Kemp. Multimodal anomaly detection for assistive robots. *Auton Robot*, 43(3), 2019.

[Pereira e Silveirat, 2018] J. Pereira e M. Silveirat. Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In *Proc. ICMLA*, 2018.

[Selcuk, 2017] S. Selcuk. Predictive maintenance, its implementation and latest trends. *Proc Inst Mech Eng B J Eng Manuf*, 231(9), 2017.

[Soelch *et al.*, 2016] M. Soelch, J. Bayer, M. Ludersdorfer, e P. van der Smagt. Variational inference for on-line anomaly detection in high-dimensional time series. In *Proc. ICML Anomaly Detection Workshop*, 2016.

[Zhang *et al.*, 2018] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, e N. Chawla. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. arxiv.org/abs/1811.08055, 2018.