Bringing Intelligence to Cyber Physical Systems via Compression and Quantization Techniques for Anomaly Detection in Industry 4.0

Dario Bruneo, Fabrizio De Vita

Department of Engineering, University of Messina, Italy, {dbruneo, fdevita}@unime.it

The advent of Cyber Physical Systems (CPS) technology has significantly changed the interaction with the world and systems in general. Indeed, the use of this technology had a strong impact also in the industrial context, leading to the birth of a new paradigm usually called Industry 4.0. In such a context, hundreds of devices with sensing/actuating capabilities communicate between them and with the surrounding environment, actively cooperating to perform one or more tasks. In particular, anomaly detection became a crucial topic in the industrial context in order to reduce unwanted (or unplanned) maintenance and avoid potentially dangerous conditions that can compromise a system operability.

Machine learning is a very important component for bringing intelligence to CPS representing one of the main building blocks for the realization of a new category of systems called Intelligent Cyber Physical Systems (ICPS). In this sense, the majority of the anomaly detection techniques proposed in the literature today involves the the Artificial Intelligence (AI) that provides the tools for the diagnosis of a system "health" state by analyzing sets of sensor of various nature (e.g., vibration, current, temperature, noise, etc.) typically in the form of multivariate time series. Through the analysis of these signals a machine learning model can learn which are those conditions in which a system properly works and then exploit this knowledge to detect abnormal (or anomalous) behaviors.

Traditionally, Cloud computing has been the leading technology for the realization of intelligent applications by providing the infrastructure and the services fundamental for the data collection, management, and processing. However, industrial applications can have very strict requirements in terms of security, latency and connection stability that the Cloud is not able to manage and which make it unsuitable. For all these reasons, in the recent period we are we are witnessing a shift of the computation towards the Edge to match the above mentioned requirements. Of course, moving the computation "closer" to the data comes with a set of challenges which are mainly related to the hardware constraints of Edge devices, that pose significant limitations on the tasks that can be executed.

If on the one hand the AI allows to accomplish very complex tasks, on the other the limitations of Edge devices do not allow to run too computationally onerous algorithms such as Deep Neural Networks (DNNs) whose huge number of parameters could not fit the hardware resources.

Modern solutions like compression and quantization techniques allow to alleviate this problem by reducing the complexity of a machine learning model such that it can fit the hardware constraints of an Edge device, while keeping its performance. Compression is a technique that aims to reduce the number of weights by "condensating" them into a limited set of clusters, while maintaining the original number precision. On the other hand, the quantization allows to keep the same amount of weights of the original model, but sensibly reduces their memory occupation by reprojecting them on a discrete interval space (typically with a 8 bit resolution).

In this work, we propose an anomaly detection algorithm based on Echo State Networks (ESNs) for the analysis of multivariate time series data. Exploiting compression and quantization techniques, we were able to deploy the algorithm directly on a microcontroller of the STM32 family, thus enabling a real time monitoring of an industrial plant. Experimental results show that the proposed application is able to correctly detect the occurrence of anomalies in the system, thus demonstrating the feasibility of the proposed approach for the realization of smart Edge monitoring platforms.