Control System Data and Meta Data at ESS

Andersson, Per; Larsson, Jan Eric; Rathsman, Karin

2021

Document Version:
Publisher's PDF, also known as Version of record

Control System Data and Meta Data at ESS

Per Andersson, Jan Eric Larsson, and Karin Rathsman
January 25, 2021

Summary report from Work Package 2 (WP2) of the ESS Control System Data Lab (ESS-CSDL) pre-study

Motivation

The European Spallation Source (ESS) will open in 2023. Here, we view it as a machine supplying neutrons to facilitate experiments performed by international scientists. The ESS machine is made up of many diverse systems, including, for example, water cooling, vacuum, power distribution, timing systems, information technology, networking, microwaves, and cryogenics. With about 100 000 devices and 1.6 million process values in the ESS control system, the set requirement of 95 % availability will be challenging. This is a strong motivation to explore machine learning (ML) technologies to improve efficient and reliable operations.

Process alarms are the key messages from the machine through its control system to the human operators indicating that there is a need to respond. Proper alarms should be actionable, provide operators with insights and situational awareness, and provide guidance towards remedial actions. Alas, existing alarm systems often fail these goals. Alarm systems may be badly tuned, may not distinguish relevance of alarms in different machine modes (stopped, start-up, running, etc.), and may cause information overload on operators. These process alarms are therefore our main target area of ML in this project.

Summary

It is clear that the complexity and richness of data in the ESS control system makes it an ideal source of data for experimenting, verifying, and validating ML on top of control system data, targeting their future use in a wide range of industrial environments. The data and meta data in the ESS system is very typical for a modern, large-scale, well-managed industrial processes. There are proper naming conventions, signals are actually following the naming conventions, and data is stored in databases and archives. This means that, technically, it can be shared with other organizations.

Required meta data also exists, is well documented, and can be shared. However, it is currently stored and managed in several different systems, and not available in a single, machine-readable general process model. This is the typical situation for almost all modern industrial processes.

Our primary conclusion is that once the different systems in ESS are designed, built, and put in operation, it will be possible to share both data and meta data with other users, for example, in research projects and commercial industrial projects.

Most algorithms and methods based on AI and machine learning will need or benefit from having a machine-readable general process model. Our second conclusion is that based on existing ESS information it is viable to fill any gaps in such a model, in part by applying ML to that sub-problem.

Given differences between machines it can be questioned if sharing data between facilities with a target to build general algorithms is at all meaningful. Our findings clearly indicate that it is, similarities of control systems structure and basic process patterns completely dominate any

specifics of the individual machine. The ESS control system seems to be an ideal source of data for experimenting, verifying, and validating machine learning and any other interesting algorithms on top of control system data, with the target to their future use in a wide range of industrial environments. The data and meta data in the ESS system is very typical for a modern, large-scale, well-managed industrial processes and industrial control systems are indeed very similar.

Data and Meta Data

The main target of the overall ESS Control System Data Lab (ESS-CSDL) pre-study is to investigate the possibilities and advantages of sharing ESS control system data with industrial companies and different research organizations. This demands both that it is possible to find good methods for open-source sharing, the actual availability and value of data, and the practical legal questions of sharing data from ESS and its different groups. The general question of availability, legality, and open source possibilities are described in Work Package 1. The analysis of availability, technical value, and usefulness belongs here. Finally, the legal/licensing requirements are described in Work Package 3.

Here we focus on the meta data required to build a data lab from control system data. To use data in any application, the data needs to be interpreted in a *context.* Typically, that context is a computer model of the process that generated the data. Most early successes in ML, for example image recognition, are based on data that has a simple, regular inherent structure. The elements of an image are numerical representations (pixels) of light and color in a specific, spatial part of the image. To process such data we really only need to know how each pixel maps to a location (coordinate) in the image, something that is already standardized in any image format since it is obviously required to display the image correctly.

The target objectives of our approach to share data from the integrated control system at ESS are that:

- The data should be usable without deep knowledge of the ESS specifics, including:
  - The physical process from which the control system data is derived
  - The control system itself and its configuration
- The methods used should be generally applicable to other industrial processes

The nature of control system data is a large volume of loosely related time-series of disparate values and statuses obtained from sensors on the underlaying physical process. This is true of both process values (analog measurements and/or status indications from the machine) and of the process alarms and is vastly different from the well-organized nature of pixel data from an image.

We defined the different types of meta data and summarize them as follows:

Pointwise meta data: This information exists within the control system, is well organized and can easily be exported as annotations to the indidual data items from the control system.
Operational Modes and States: The concepts of mode and state are well defined and documented at ESS. This actually puts ESS in the forefront of the industry – the concepts are normally used without much of a formal definition. Still, mode and state are not explicitly represented in the control system and cannot be exported. Instead this is meta data that need to be mined from the primary control system data. For modes, it is relatively easy as modes in principle reflect the commanded settings in the control system. For states, we propose an ML based follow up project to derive proper information.
Meta data model: this is not something that normally exists in a control system, or even at all as an aggregated, machine readable description. On the other hand, the usability of control system data is

tightly tied to the availability to some form of meta data model. We have investigated the various data available and proposed a way to represent a meta data model. At ESS, all the relevant, raw information to derive such a model is available and well structured.

Roadmap

This project was a pre-study to prove viability of using the ESS control system data to benefit Industry 4.0 development, to outline a roadmap to exploit those possibilities and to eventually serve the industrial needs in cooperation with future partners. The proposals have been selected due to their general applicability across industrial control systems. We believe that the proposals will demonstrate the true value of ML from control system data.

Creating a Digital Twin Instance by Machine Learning: Currently, ESS has a good situation concerning availability and storing of both primary signals and alarm data, as well as different kinds of meta data, such as documentations and descriptions. However, this data is not available in a single, general, machine-readable format. This project would derive a formal description – effectively a digital twin instance – with functional blocks, tied to their physical and logical process objects, and with the location of processes and objects. It would use an interchange format using standards such as RDF or JSON, and define methods for version control of these descriptions to match process data over time. This type of model could be built with limited resources. We propose to use machine learning to build this model, by converting traditional P&I diagrams and existing data bases into the proposed description. *This is the main vehicle in the effort to interpret control system data without deep machine/process knowledge, thus making sharing of the data truly meaningful.*

ML-Based Operation State Identification: The operational state is the current state of single process variables as well as of components and of the entire process. Typical (global) process states are off, stand-by, starting, normal operation, shutdown, emergency shutdown, etc. They differ from operational modes in that modes are the situations intended by the operator, while states are the actual situations the process end up in. Alarm management algorithms can and should be used to suppress alarms that are irrelevant in a certain state, but it is very important that the state is robustly identified, so that the alarm suppression is correctly done. We propose to use machine learning methods to monitor process variables and identify the operational state, and to detect when the operational state changes. *This has the potential to relieve true, state based alarming from the burden of manual configuration.*

ML-Based Fault Detection: Process faults are a threat to both operation and safety, and need to be detected quickly, and identified quickly and robustly, in order to enable rapid and correct remedial actions. Faults are typically detected via the alarm system, but both detection and identification can potentially be improved by looking at a larger number of process variables. We propose to use machine learning methods to monitor both process variables and existing alarms, to detect and identify process faults. *This goes beyond autotuned and dynamically computed alarm limits by learning fault situations from clusters of process variables and learning the typical actions to remedy those faults.*

ML-Based Root-Cause Analysis of Alarm Cascades: Faults in processes often lead to consequential faults, and if these in turn cause faults, the result can be a large chain of faults (and thereby alarms), a so-called alarm cascade. These cause a large amount of alarms and can hide the important alarms and mentally overload the operator, especially in critical fault situations. The "shapes" of alarm cascades are determined by the inherent causality of the process, and the fault chains are often similar. We propose to use machine learning methods to identify complex fault chains and replacing

them with single root cause alarms, to reduce the alarm load and improve the identification of complex fault situations. *This advances the most difficult problem in alarm analysis – finding the root cause – by enhancing existing algorithms, e.g., by automatically identifying causal chans of events during normal operations and using that knowledge in fault situations.*

Model-Based Root-Cause Analysis of Alarm Cascades: The problem of alarm cascades caused by a large number of consequential faults is one of the most difficult alarm problems. It creates alarm overload situations during complex faults situations, that is, exactly when the alarm system is needed the most.

Model-based, symbolic AI methods can solve the problems of alarm cascades efficiently and correctly. Such algorithms have been deployed successfully for different complex systems, such as, for example, electric power grids, and such a system has been monitoring the Swedish National Grid since 2007. We propose to develop models for ESS systems using the same methodology, and applying AI-based algorithms for real-time root cause analysis. We believe that the methodology can be adapted to most ESS systems. These algorithms have been shown to reduce the amount of alarms by factors of over 99 % in, for example, power grids. *This project would directly demonstrate the usability of the digital twin instance from step one above as the required model could be directly derived from that instance.*

Full Report

This document is a summary of the full report from Work Package 2: “Per Andersson, Jan Eric Larsson, and Karin Rathsman. ESS control system data lab - work package 2 final report. Technical report, ESS-3216740, 2021.”