Industrial Machine Learning Model Validation

1 Introduction

Maillefer Extrusion develops industrial machinery for producing power cables and fiber optics. These devices are increasing smart and networked together, in what is known as the fourth industrial revolution, or Industry 4.0. With smart devices generating and consuming vast amounts of data, machine learning (ML) and big data solutions are rapid being integrated with more traditional industrial equipment.

We have recently launched a measurement device that scans product surfaces for defects, using neural networks to analyze the measurements. As with any ML system using neural networks, the amount of data needed to train the networks and validate how well they work is huge.

Constantly collecting new data to boost the size and quality of the training set enables incrementally improving the performance of a neural network. For a production system, this directly translate to more customer value. Before a newly trained neural network, or any other ML model, is deployed to a live environment, it must be validated. The validation step involves testing the new model against real-world input data and analyzing the results to measure the model's accuracy, as well as ensure no regressions happen. As the datasets involved are large, the validation process should be automated, which is the topic of this project.

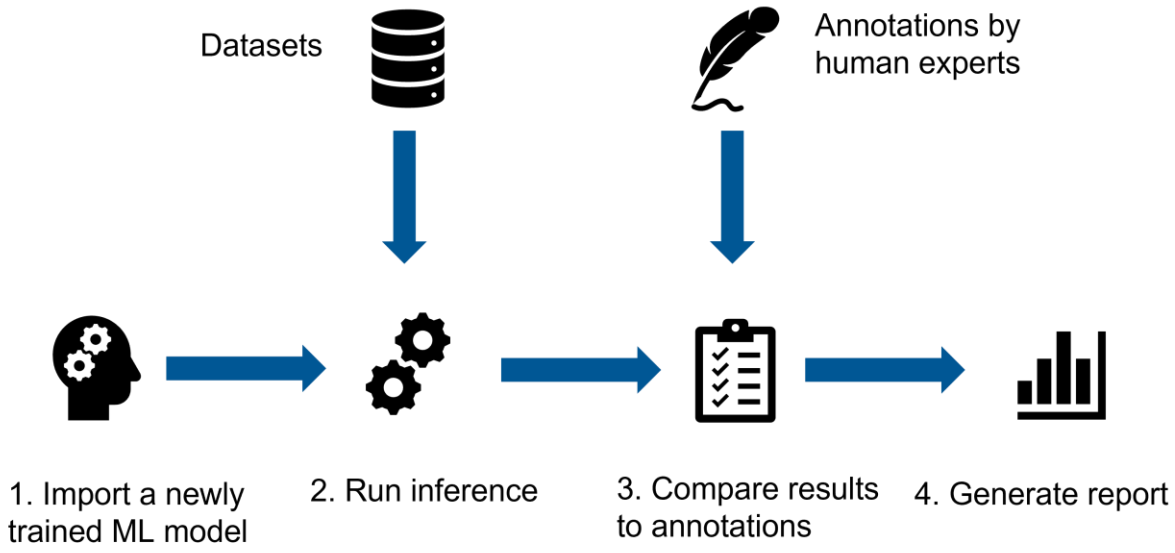
This project presents an opportunity to acquire experience in how machine learning can be used in industrial production environments.

2 Project Goal

The goal of this project is to implement a system for validating how well a new machine learning model performs by using the model on pre-existing datasets and comparing the outputs to expected results, defined by human experts. This process is shown in the following figure, with the parts implemented in this project marked as numbered items:

Maillefer Extrusion Oy
Ensimmäinen savu, P.O. BOX 176
FI-01511 Vantaa, Finland
Phone +358 9 88 665 600, Fax +358 9 88 665 771, www.maillefer.net
VAT No FI16577533, Domicile: Helsinki, Finland

MAILLEFER



The system user interface can be anything from a command line interface to a custom webapp, or built using pre-existing data science tools, such as Apache Spark and Jupyter. Ideally the UI is easy to use, for example a drag & drop the model onto a web app, select datasets to use and later explore the results in the same web app, but the UI is not a priority feature.

The reporting phase presents an opportunity for the team to experiment and innovate, once the core system is implemented. A basic report contains simple statistics, like how many items in a dataset were correctly classified by the model, versus how many were incorrectly classified. Advanced reporting features can calculate more detailed statistics or apply big data techniques to identify common features in misclassified items.

Understanding exactly how the evaluated model behaves is the ultimate objective of this project, and valuable for improving the model itself as well as the training process.

3 Technologies

The suggested implementation approach would be a cloud native system using one of the big three cloud providers, as all of them offer ready-to-use services suitable for implementing parts of the system, such as batch inference for datasets. Internally we use Microsoft Azure and can offer support with the platform.

We are also happy to accept proposals for implementing the system using components that can run locally as well as in the cloud, such as Docker containers. Please list relevant tools and technologies you know in your team CV!

Most cloud environments and machine learning tools use Python as the native language and a working knowledge of Python will be beneficial for this project. Familiarity with C/C++/C# can also help.

4 Requirements for the Students

Willingness to learn the project domain and new technologies is the primary requirements for this project. The resulting tools are intended for internal production use and further development, so a quality-oriented software craftmanship mindset is highly appreciated.

Prior knowledge of machine learning is not required. We will supply all the needed ML components and guide the team in their use. However, we consider experience and interest in ML and/or big data extra merits in the team's CV. Additionally, any prior experience working with distributed and parallel systems is a plus.

We expect the team to follow course instructions for applying Scrum to the development process, as well as established best practices of software development, such as using Git for version control, implementing standards for code quality, peer reviewing all deliverables etc. A member of our team has acted as a coach on this course for several years and can offer tips on what practices dev teams have found valuable.

5 Legal Issues

Intellectual Property Rights (IPR)

The resulting intellectual property rights to all the results will belong to the client.

Non-disclosure Agreement (NDA)

Signing the NDA included in the Aalto contract is required. This is mostly for convenience: to be able to work at the client's premises and handle the client's real industrial data.

6 Client

Short Introduction

Maillefer Extrusion is a Finnish-Swiss engineering company with main offices at Vantaa employing 400 people. Internationally we have second site in Lausanne, Switzerland, and branch offices and representatives in 8 countries. Maillefer has a decades long tradition of developing specialist software for the field, for example analysis and simulation tools.

What do we offer?

We hope to arrange a social kickoff event (COVID situation allowing) with food and drinks at our offices in Vantaa, so everyone gets to know each other. During the project, all sprint meetings and demos can be held at our offices. Working space can also be arranged for other co-located working sessions if the team wants to work on the premises.

Who are we?

The development team will work with our R&D team, many of whom are Aalto University alumni from various fields, mechanical & chemical engineering, automation and computer science.

Product Owner, Process Development Manager
Janne Harjuhahto
janne.harjuhahto@maillefer.net
+358 50 532 7531

Data Scientist
Jussi Hanhirova
jussi.hanhirova@maillefer.net

7 Additional Information

There exists a possibility of continuing to work with this and similar projects after the course under internships or junior developer positions.