THIRY Zachari
RAMON Matthieu
SALON Benjamin

MALIS
Fall 2021

Machine learning in Industry: Using ML regression models for RUL prediction in predictive maintenance.

Motivation:

Our project consists in finding a manner by which machine maintenance can be optimized in the production industry. Most modern and high-tech machines nowadays have sensors able to record features such as machine temperature, rotation speed, or running time which are used by predictive maintenance analysts to predict tool wear and schedule maintenance.

Our goal is to propose a machine learning algorithm to predict such failures with great accuracy to comply with industry standards.

In practice, we aim at estimating the Remaining Useful Life (RUL) of consumable machine tools using machine and production data.

Method:

A first look at our datasets indicate that the RUL of a tool appears to be linearly linked to its age (running time), and slightly differs according to other factors such as torque, accuracy, temperature and more. We thus believe our output can be estimated using a regression model linear with respect to sub features computed from the features values.

For a first approach, we intend to implement such a model using our datasets and predict the time at which the tool will fail, in other terms the time at which our predicted RUL function will cross the abscissas axis.

Intended Experiments:

Before actually launching the learning process, we plan to better understand the correlation between the features and tool wear by creating physically meaningful sub-features and analyzing their correlation with the tools RUL. The first step will be to manually search some first correlations to determine which hypothesis class and which features or computed parameters to use in regression. We have for instance already notice the apparently linear dependance regarding the usage time.

We will then try different regression models with different hypothesis classes and features to use and compare their accuracy in predicting the correct time. We can for a first approach think of a linear regression regarding only the time used by the machine and gradually add some raw or computed features that we think play a role in the RUL.

We intend to divide our data in three separate sets following a 60 - 10 - 30 which leaves us enough data for the training and a test data set sufficiently large to obtain meaningful validations of the models.

It's not excluded that if we find results fast enough, we look at different machine learning model to answer the question from a different angle such as SVM Classifier.

Dataset and research:

We have selected two datasets that could be interesting for experimenting. The first dataset is the AI4I 2020 Predictive Maintenance Dataset Data Set from the *UCI Machine Learning Repository* which consists of 10000 data points with 14 features of usage sensor data of a single industrial machine. The second dataset is a *Kaggle* dataset, https://www.kaggle.com/calpara/predictive-maintenance which is a collection of telemetry sensor data and failure reports from 100 industrial machines.

We read this publication to have a first glance of what can be done in the field: *A systematic literature review of machine learning methods applied to predictive maintenance*, Thyago P. Carvalho, Fabrízzio A. A. M. N. Soares, Roberto Vita, Roberto da P. Francisco, João P. Basto, Symone G. S. Alcalá, Computers & Industrial Engineering, Volume 137,2019