Nuclear Science and Technology Symposium - SYP2019
Helsinki, Finland, 30-31 October 2019

Utilizing Machine Learning in Nuclear Power Plant Lifecycle Management

Pirkkalainen Herkko
Platom Oy
Jääkärinkatu 33, 50130 Mikkeli
herkko.pirkkalainen@platom.fi

ABSTRACT

Nuclear power plants (NPPs) have thousands of devices for collecting data with both periodic and online measurements. The devices age over time, which eventually leads to device failure. These failures are recorded and stored as statistics. Ageing is prevented and slowed down by performing maintenance on the devices. When planning the maintenance of these devices, the order of the maintenance tasks is prioritized by certain criteria.

In machine learning, a computer program is trained to make independent conclusions from the data it receives. The machine is first provided with training data, which includes a set of measurement results and possibly also man-made conclusions about the data. The machine then goes through the data using the chosen machine learning algorithms and formulates rules regarding the internal classification of the data. New data can then be provided to the machine, which will make new conclusions about the data within a certain confidence interval.

This paper is based on the author's master's thesis made for Teollisuuden Voima Oyj (TVO) in 2019 [1]. The most important tool used in the thesis was the MATLAB computing program and its machine learning modules made by MathWorks Inc. Three different cases were looked at: Detecting a feedwater pump axle seal leak, analysing the reactor pressure vessel surface level measurements, and rationalizing the prioritisation of the maintenance of nuclear power plant systems. Utilizing machine learning provided promising results in the case of the feedwater pumps and especially in the case of the surface level measurements but proved unsuitable for rationalizing the prioritisation of maintenance tasks. Basic knowledge of the different methods of machine learning was acquired, and some of them will be utilized in the future at the Olkiluoto NPP (OL).

1 INTRODUCTION

The purpose of the author's master's thesis was to find out how machine learning could be utilized in the lifecycle management of the Olkiluoto NPP. The first point of focus was to see if future malfunctions of powerplant machinery could be predicted in advance. The case studied was the early detection of the leakage of feedwater pump axle seals.

The second point of focus was to see if different machine learning algorithms could detect some previously unknown aspects of the state of the NPP by going through regular process measurement data. The case studied was the comparison and separation of four reactor pressure vessel (RPV) water surface level measurements.

The third point of focus was to see if machine learning could be utilized in the rationalization and prioritization of NPP lifecycle management projects.

The goal of TVO is that in the future the system- and device owners and heads of engineering could utilize this information. The future placement of machine learning in the lifecycle management process of the Olkiluoto NPP can be seen in figure 1.

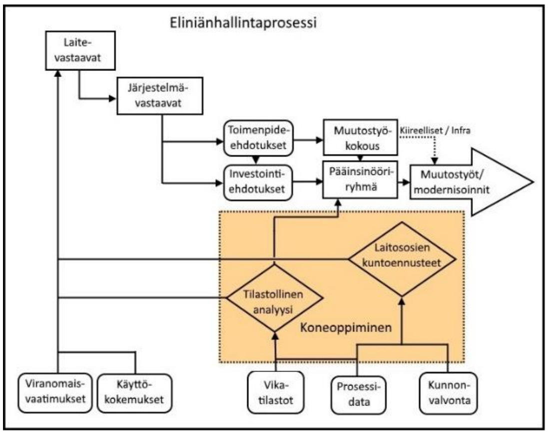

Figure 1: A simplified chart of the TVO lifecycle management process, where the orange backround signifies the future place of machine learning. (Vaaheranta, 2018) [1]

2 SOME PRINCIPLES OF MACHINE LEARNING

Machine learning can be roughly divided into three different approaches; supervised learning, unsupervised learning and reinforcement learning. These approaches can also be used together to make new combination methods. [2]

In the case of supervised learning, the machine is given a training data set with a specific activity assigned to each point of data. The activity could be, for example, that ”the car is moving forward” or “the axle bearing is overheating”. Activities are also called supervisory signals, which is where the name for the supervised learning comes from. After training, the machine is supposed to have learned the connections between the various data points and their corresponding activities. The machine should then be able to classify new data with the same activities correctly. Supervised learning is used when it is already known what the different datapoints represent in real life [2].

In the case of unsupervised learning, the machine is given training data without activity information. The machine will then try to recognize groups, patterns and connections in the data using machine learning algorithms. At the end of the training, the machine will produce data classified into certain groups or categories, but they will not have any specific activity tied to them. It is then up to the user to assign the desired activities to the data groups, which requires fundamental understanding of the original training data. These unsupervised methods are used, when the relations between the different data points are unknown and the data set is very large [2].

The general rule is that the more training data is available, the better the trained machines will become. Larger data sets will of course make the training slower, especially if the data has high dimensions, and there are distinct differences between different machine learning algorithms in the way they cope with very large data sets. Very large data sets are also prone to cause overfitting in supervised learning, if the training activities are numerically very imbalanced.

3 MACHINE LEARNING METHODS USED

Supervised learning was the method mostly used in the thesis. The MATLAB modules used were the Classification Learner (CL) and the Neural Network Pattern Recognition (NNPR). Similar work could have been done, for example, with the open source machine learning library TensorFlow made by Google Brain Team [3]. Unsupervised clustering methods were tried but, unfortunately, they yielded no practically usable results. Further orientation into the methods and tools would have been required, and therefore unsupervised learning was left out of the thesis’s scope. Some of the most common machine learning algorithms can be seen in figure 2.

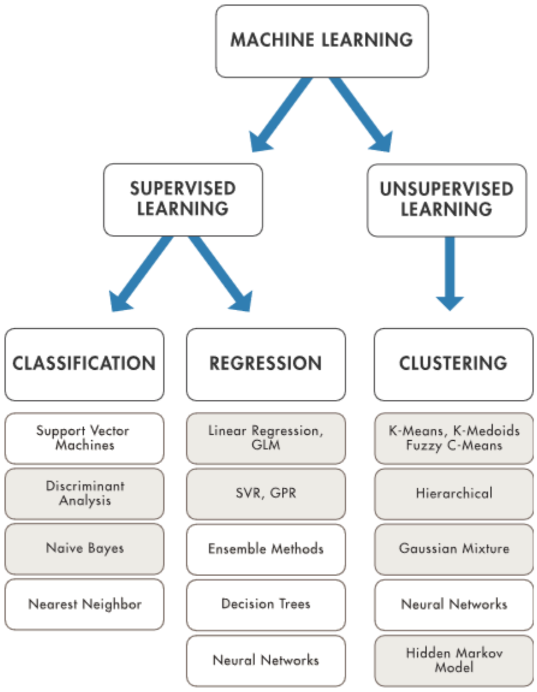

Figure 2. Some of the most common machine learning algorithms, divided into supervised and unsupervised learning. The algorithms used in the thesis have a white backround. (MATLAB, 2018) [4]

4 HANDLING THE OLKILUOTO NPP MEASUREMENT DATA

The OL1/OL2 NPP units have a large number of sensors measuring the various process variables, that are used to monitor both the state of the processes and the condition and operation of individual devices. The process variables can for example be temperature, pressure, volume flow, rotational speed or position. The measurement data of the variables is collected with two different systems. The first one gathers data with a low frequency from an extended time period (MTD), and the other one gathers high frequency data from a short time period (KTD).

The pretreatment of the data was done by making a MATLAB script collection that packed the MTD-spreadsheets into the same format that the KTD-data uses. Both datasets could then be manipulated and analyzed with the same scripts.

All measurement data will most probably have some systematic and random errors. These errors must be acknowledged and hopefully compensated for before feeding the data to the machine learning algorithms.

5 RESULTS

5.1 The case of the feedwater pump axle seals.

Three activities were created for the training data: “normal”, “slow leak” and “fast leak”. The division into the activities was made by hand after carefully examining the initial MTD-data. After an axle seal has started leaking, the time of occurrence can be narrowed down to a 2-5-hour period. The speed of the leak can also be seen.

Several trained machines were able to classify the test data with reasonable accuracy, and therefore they could be used to recognize a seal leak in its early stages. The accuracies of four different algorithms can be seen in table 1. The test data used was raw data obtained from the OL1 unit, where a new seal leak had occurred during the writing of the thesis. However, it is important to remember, that the success percentage is directly dependent on the numerical balance of the different activities. Both the test and the training data had approximately 95 % of the activities as “normal”. Therefore, even if the machines classified the test data as completely normal and didn’t notice the leak, they would still be correct 95 % of the time by default.

Table 1. The success rate of different machine learning algorithms found in the CL-module.

Machine learning algorithm	Success percentage
Classification tree (fine)	98.3 %
SVM (quadratic)	98.1 %
k-NN (fine)	98.6 %
Ensemble (forest)	98.7 %

The best classifier was a combination method of different classification trees (ensemble forest). It predicted the “large leak”-activity with 91 % accuracy and the “small leak”-activity with 61 % accuracy. The amount of training data had the smallest impact on the accuracy of the k-nearest neighbour classifier (k-NN) and the largest impact on the support vector machine (SVM) classifier. The SVM was also the slowest to train.

5.2 The case of the RPV water surface level measurement

Four activities were created, each representing one of the four water level measurements. The test data and the training data were taken from separate days. Several of the trained machines were able to identify and separate the signals from each other with good accuracy, showing that measurement signal chains that are identical on paper still have distinct features. The amount of the training data had a large impact especially on the neural networks, which require a lot of it. A large training data set however slows down the training process, and for example discriminant analysis’ and SVMs can’t handle very large data sets with high dimensions. The results of the training can be seen in table 2. The best algorithm for this data type and size was again the ensemble forest.

Table 2. The success rate of different machine learning algorithms found in the CL-module and the NNPR-tool.

Machine learning algorithm	Success percentage
Classification tree (fine)	94.8 %
SVM (quadratic)	94.9 %
k-NN (fine)	94.6 %
Ensemble (forest)	99.9 %
Neural network	94.7 %

5.3 The prioritization of NPP lifecycle management projects

The systems of OL1/OL2 NPP units are currently classified by three lifecycle management metrics; importance to safety, importance to plant economics and the overall aging factor. The maintenance and renovation of the systems is partly based on those metrics. There are several hundred underlying metrics that can be used to calculate the numeric values for the three lifecycle management metrics, which is currently done by a MATLAB script library made by TVO.

The underlying metrics were given to multiple different machine learning algorithms. The best algorithm in this case was the k-NN classifier with an accuracy of 99.7 % when classifying the state of OL1/OL2 systems, as compared to the TVO script library “accuracy”. All the trained machines reached an accuracy of over 98 %.

Machine learning could be used to accurately prioritize NPP lifecycle management projects in the future, but a 100 % success rate would require a more thorough fine tuning of the algorithms. The application of machine learning in this case was not practical, because the current TVO script library can prioritize the systems instantly with perfect accuracy. Machine learning could become relevant in this case, if clearly more underlying data was used. The current MATLAB scripts would then have to be redesigned, which could possibly take much more time compared to using a suitable machine learning algorithm.

6 DEVELOPMENT IDEAS

6.1 Machine learning & mechanical devices

In the case of the feedwater pump seal leaks, the trained machines could be significantly improved by tweaking their individual variables and further improving the preprocessing of the training data. New, possibly relevant process variables could also be added to the mix. The best of these improved machines could be used to alert the system operators of a seal leak in a very early stage, without giving false alarms.

With further development, a machine could be trained to predict a seal leak in advance based on the state and future trend of the process variables. One future possibility would be to implement a trend plotter, which would send out alarms if the current trend of the process variables would lead to an eventual leak. The idea is shown in figure 3.

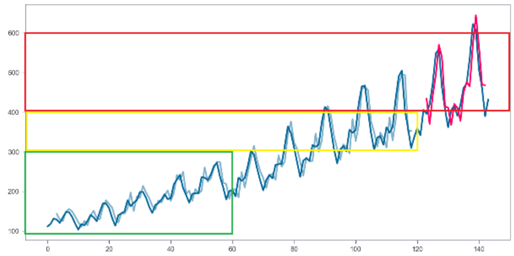

Figure 3. An illustration of a possible model predicting the probability of axle seal leaks in feedwater pumps. The blue line is the projected trend of a process variable, and the different colored boxes indicate an increased risk of a leak appearing.

The same principles could be applied to all mechanical devices in the NPP. Various maintenance tasks of pumps, valves, motors etc. could then be done at an optimal time with optimal efficiency.

6.2 Machine learning & electronic components

In the case of analyzing the RPV water surface level measurements, the individual signals could be further dissected in order to find out what causes the disparities in them. The maintenance data of the various components in the measurement signal chains could be combined with the current data to find out how each component might affect the final signal output.

When the next maintenance or replacement of a component in the measurement chain takes place, the difference between the old and the new measurement signal could be analyzed to find out how the components affected the signal. In the future, with a successful analysis of signals, it could be possible to predict which component of the measurement chain is reaching the point of failure.

The same principles could again be applied to other signal chains and electronical components in the NPP.

6.3 Other ideas for improvement

In the future, it would in theory be possible to gather every single variable found in the MTD and KTD process measurement systems and combine all of it into a single database in a format suitable for machine learning. One could then also gather every single plant event from history and add them to the database (Big Data). Utilizing machine learning methods to analyze this combined, massive data set might reveal currently unknown cause-effect relationships between the different NPP systems. This would of course demand a vast amount of resources.

Machine learning methods could also be utilized by NPP personnel not educated in the subject to improve their work efficiency in data analysis for example, if the new machine learning tools are user friendly and inexpensive to use.

REFERENCES

[1] H. Pirkkalainen, ”Koneoppimisen Hyödyntäminen Ydinvoimalaitoksen Eliniänhallinnassa”, Master’s Thesis, 2019. Available at: http://lutpub.lut.fi/handle/10024/159299

[2] S. Shalev-Shwartz, S. Ben-David, “Understanding Machine Learning”, 1st edition, Cambridge University Press., UK, 2014.

[3] A. Géron, “Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”, 1st edition, O’Reilly Media, 2017.

[4] “Machine Learning in MATLAB”, [internet-site]. [referenced 29.6.2019], MathWorks Inc. Available at: https://se.mathworks.com/help/stats/machine-learning-in-matlab.