Concept, Implementation and Evaluation of Machine Learning Models for Anomaly Detection of Hydraulic Systems on Basis of Industrial Data of a Machine Tool

The industrial world is in the middle of a new revolution, termed as Industry 4.0, which is characterized by its reliance on the use of cyber physical systems. These are able to communicate with each other and make autonomous, decentralized decisions to increase industrial efficiency, safety and transparency. [1] In this context technologies such as cloud computing, sensor networks and other data services merge into the so-called Internet of Things. They provide an ideal connectivity between industrial processes, workforce and users to exchange information in order to better manage on-time production and thus enhance profits. [2] Thereby a wide range of information, such as process, plant or maintenance data, becomes available for analysis which provides the opportunity to employ data-driven condition monitoring for a condition-based maintenance [3]. With data-driven condition monitoring, possible performance degradation or defects of production machinery are detected to improve the planning and management of maintenance processes. In this way, production delays due to equipment failure are minimized. [2] In addition, the condition monitoring approach offers the possibility of making maximum use of the available remaining useful life of machinery components by not replacing intact equipment through preventive maintenance measures [3].

One example of such systems that can benefit from the use of condition-based maintenance are hydraulic aggregates. These are widely used in modern machinery for their numerous advantages, such as a fast response, significant load stiffness, large power density, and superior stability [4]. Some operation fields of hydraulic systems are mining, pulp, paper, agriculture, construction, and aerospace [5]. Often hydraulic systems thereby compose the central component of engineering equipment, such as control and power transmission systems. These machines can suffer from environmental influences such as sunshine or dust particles and unstable working conditions such as a high load or severe impact. [4] Hence, faults in hydraulic systems are difficult to avoid and will result in a deteriorated performance of the equipment [6]. Hydraulic pumps can suffer of wear for example because of improper fluid selection, cavitation or overload. A common failure mode for actuators is leakage, which can be caused by deterioration of seals. Valve faults can include surface wear and solenoid faults. One cause for the surface wear and deterioration of seals is a contamination of the used

hydraulic oil. Other issues are abrupt failures such as blocked actuators, spools or nozzles in the valves. [5]
For this reason, it is important to monitor and detect failures at an early stage so that maintenance can be scheduled proactively when necessary in order to increase the availability of the machinery. To that end, the condition of the equipment showing signs of loss of performance or future failures, must trigger the maintenance so repair can be completed before the performance falls below the optimum level or the machine fails. [7] However, an appropriate fault diagnosis for hydraulic systems remains a challenge. In comparison to mechanical or electrical installations, faults in hydraulic systems employed in engineering equipment are more hidden and unclear. Consequently, it is challenging to obtain information about the fault and the relationship between a fault characterization and its cause. Hence the research of key technologies and methods for achieving hydraulic fault diagnosis is an important task. [4]
Apart from these rather specific problems concerning the condition monitoring of hydraulic systems, there are also more general difficulties in monitoring machines such as a lack of understanding of the monitored equipment and mainly the absence of data on failure cases and the evolution of the degradation. These kinds of data can only be collected if the machine equipped with sensors runs until an anomaly or failure occurs, which is unwanted by the industry. Additionally, gathering fully representative datasets for the analysis is difficult due to the varied types of failure and operating conditions. For these reasons test rigs are usually used to simulate anomalies and failures. However these systems have the disadvantage of only offering simulated failures of several components which are already widely studied. [7]

It turns out that condition monitoring for early fault detection in hydraulic systems can make an important part in improving their reliability by detecting and avoiding failures at an early stage. In this sense, the main contribution of this thesis is the elaboration and evaluation of two components of the condition monitoring approach, namely the anomaly detection and isolation, for a hydraulic aggregate.
The selected methods for this task are applied to the data of a real hydraulic aggregate installed in a machine tool. There the hydraulic system takes over tasks such as clamping workpieces and locking tools and machining tables. A failure of the installed hydraulic aggregate therefore has serious consequences for the correct functioning of the entire machine and would normally result in a shutdown of the machine tool. This would in consequence lead to a loss of production capacity, which in turn would cause an economic damage. To avoid this, a monitoring system for hydraulic systems is proposed in this work, capable of reliably detecting anomalies caused by equipment failures. Furthermore, the sensors whose measurements deviate from the normal sate are determined to identify the source of the anomaly. In the best case, this procedure would allow to repair or replace the affected parts before the the hydraulic system fails and thus avoiding a machine failure. For this purpose,

different methods, which build on the data-driven approach are investigated and compared with each other. This approach offers the advantage, that the different models can be trained with historical data and hence no exact understanding of the partly very complex hydraulic system is needed.

To provide the necessary data basis, the hydraulic aggregate was retrofitted with several sensors in advance, whose measured values were recorded with additional information from the machine control system. The recorded data from the real operation of the system are analyzed and processed to be used for training and evaluation of the examined methods for condition monitoring. In this process, the data sources with the most information content, which in this case are mainly pressure and temperature sensors, are identified and composed. One challenge is that the data was recorded during the regular operation of the machine tool. For this reason, no labelled data is available for deliberately induced operating states or simulated fault cases. This limits the application of the frequently used supervised learning methods for condition monitoring.

To avoid this limitation, four different procedures that do not rely on the availability of labelled data are evaluated in this thesis to realize a condition monitoring approach for the considered hydraulic aggregate. Three methods built up on the novelty detection principle: the cluster based, the one-class nearest neighbor and the one-class support vector machine approach. These models are trained with the data representing the faultless operation of the hydraulic aggregate, which is available in sufficient quantity. This enables the models in a subsequent step to detect deviations from this state, for example because of an equipment fault. Thereby, the examined methods are based on different approaches to determine whether a new data point represents a state similar to the error-free state used in training or whether it differs from the error-free sate. In addition to the detection of anomalies, two of the three methods are extended with an approach for isolating the source of the anomaly. In this step the sensors, whose measured values deviate most form the trained normal state shall be identified to support the maintenance staff in finding the anomaly source.

This anomaly isolation is illustrated as an example for the case of the one-class nearest neighbor method in figure 1. In this procedure, the distance between a new data point to its nearest neighbors in the training dataset is determined. At the same step, the relative share of the considered sensors in this distance is also taken into account to determine which measured value is most responsible for the deviation form the normal state. This allows to identify the corresponding process parameter that deviates most from its normal value.

In addition to the methods relying on the novelty detection approach a time series-based approach is investigated. This procedure is based on the use of a time series model to map the fault-free state of the aggregate and to detect possible faults by comparing the model output with the measured values of the system.

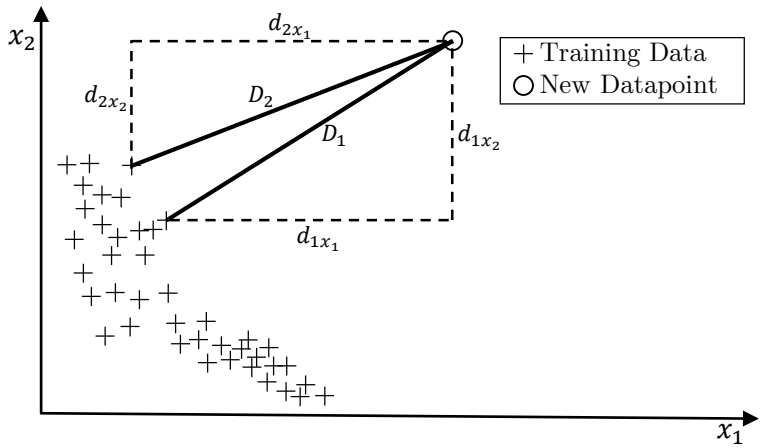

Figure 1: Anomaly Isolation of the One-Class Nearest Neighbor Approach

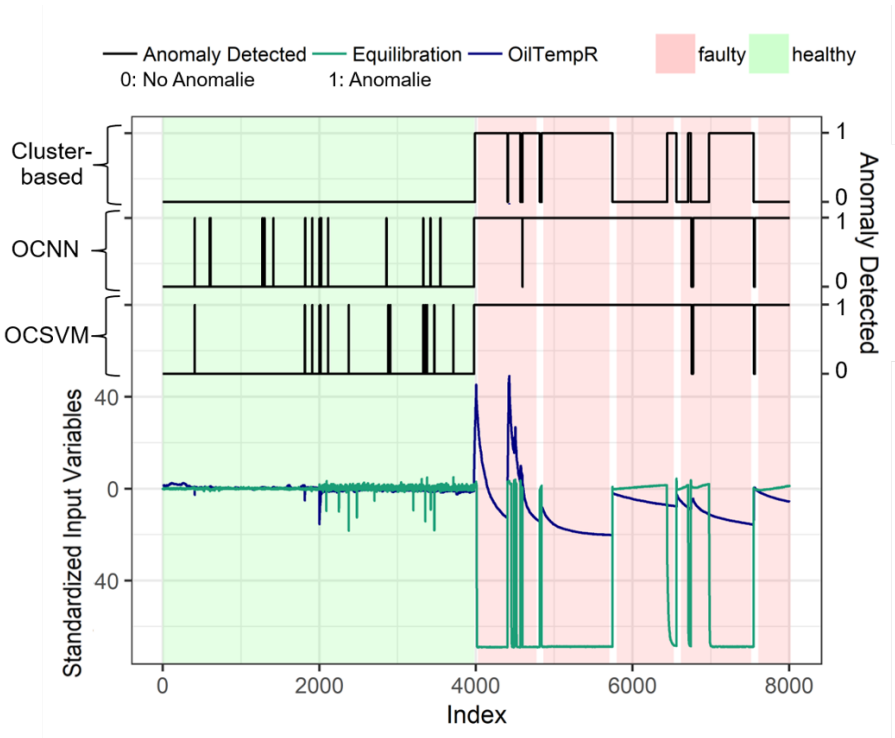

Figure 2: Results in Detecting Anomalies on the Test Dataset Consisting of Data on Normal Operation and Five Successive Fault Cases

To check the performance of the four proposed procedures, they are applied to a defined test dataset consisting about 8000 individual data points representing the error-free and faulty states in equal shares. The latter are representations of five real occurred deviations of the aggregate from its normal state as a consequence of e.g. an abnormally high oil temperature. The corresponding data set is shown in figure 2 with the standardized values for a pressure (*Equilibration*) and a temperature sensor (*OilTempR*). The first half of the data represents the data for error-free operation of the unit, while the data from index 4000 represent the five error cases. Various versions of the examined procedures with different hyper parameters are tested in order to find the respective setting with which the best possible results can be achieved. Thereby, different metrics, such as the accuracy in classifying the states but also the required processing time, are used to evaluate and compare the individual performances.

The first examined method based on the novelty detection principle, the cluster-based approach, is an easy to implement method that can classify the test data relatively quickly. However it only achieves an accuracy of 79.96 % when classifying the test data. Ideally, the model would detect all errors (starting at index 4000 in figure 2) as anomalies. However, there are some connected areas where a fault is present, but the model does not detect any anomaly. The greatest weaknesses of the cluster-based approach are *fault 3* (around index 6000) and *fault 5* (index 7600 and higher). These incorrectly classified cases are faults whose data deviated only slightly from that of the normal state. However, since the detection of even these small deviations is part of the successful implementation of an anomaly detection, this method is not a reliable method for this particular application.

In contrast, the procedure based on the one-class nearest neighbor approach (OCNN) achieves a good accuracy of 99.23 % in detecting anomalies, which is significantly better than the performance of the cluster based approach. Hence, a large part of the data from the test dataset is classified correctly. Even the faults, which are characterized by a rather small deviation from the normal state, are identified as such as can be seen in figure 2. Only some individual, data points representing the error-free state are incorrectly recognized. Furthermore, using this method, a concept could be implemented successfully to isolate the source of an anomaly. The disadvantage of this approach is the significantly longer processing time for classifying new data points compared to the other approaches.

The best results in terms of the accuracy in detecting anomalies and the time required to classify the test dataset is obtained by the one-class support vector machine approach (OCSVM) with an accuracy of 99.50 %. This method exceeds the performance of the previous two in terms of both accuracy, which is only slightly better than for the one-class nearest neighbor approach, and the time required to classify the data. Therefore this approach is perfectly suited for a fast and reliable detection of anomalies even if these differ only slightly

from the normal state. However, no subsequent isolation of the anomaly could be implemented for this approach, which unfortunately implies that this method alone cannot be used for a complete condition monitoring of the hydraulic system.

In addition to the three methods based on the novelty detection concept, the time series based approach is evaluated. In this one an ARIMAX model is used to observe two sensors of the hydraulic aggregate simultaneously. It is demonstrated that at least with one time series model, observing te relation between oil pressure and temperature, relatively good results can be achieved in the detection of anomalies. However, it was not yet possible to prove that the approach can reliably detect also small deviations from the normal state. For this reason, a fully dependable anomaly detection using this approach cannot be guaranteed in the examined form. Nevertheless, this approach has the advantage, that it provides an anomaly isolation without additional adaptations. If one of the models detects an anomaly, the cause can be traced back to one of the two time series used.

In summary, it can be concluded that both the one-class nearest neighbor and the one-class support vector machine approach, each with achieved accuracies in the classification of data on fault-free and faulty states of the hydraulic aggregate of over 99 %, are best suited to detect anomalies. Even, faults that differ only slightly from the normal state are reliably detected by these two methods. In addition to the good results in detecting anomalies, the one-class nearest neighbor based approach can also convince in the isolation of anomalies. Thereby, the input variables of the model, which had the major share in the detection of the anomalies, are identified to support the maintenance staff in the search for the fault source. Due to the overall good results, the one-class nearest neighbor based approach is the most suitable of all investigated methods for the realization of a condition monitoring of the hydraulic system.

1 Bibliography

[1] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial internet of things (iiot): An analysis framework", *Computers in Industry*, vol. 101, pp. 1–12, 2018. DOI: 10.1016/j.compind.2018.04.015.

[2] K. Shahzad and M. O'Nils, "Condition monitoring in industry 4.0-design challenges and possibilities: A case study", in *2018 Workshop on Metrology for Industry 4.0 and IoT*, (Brescia), IEEE, 2018, pp. 101–106. DOI: 10.1109/METROI4.2018.8428306.

[3] N. Fabritz and I. Weiß, "Fehlerdiagnose an ventilen, Herausforderungen, ergebnisse und zukünftige möglichkeiten", in *Automation Symposium 2018*, September, 2018, pp. 39–52.

[4] J. Dai, J. Tang, S. Huang, and Y. Wang, "Signal-based intelligent hydraulic fault diagnosis methods: Review and prospects", *Chinese Journal of Mechanical Engineering*, vol. 32, no. 1, 2019. DOI: 10.1186/s10033-019-0388-9.

[5] P. C. Palavicino, W. Lee, B. Sarlioglu, and T. Minav, "On faults and diagnostics in electrified hydraulic actuator systems", in *2018 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe)*, 2018, pp. 1–9.

[6] Z. Yao, Y. Yu, and J. Yao, "Artificial neural network–based internal leakage fault detection for hydraulic actuators: An experimental investigation", *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 232, no. 4, pp. 369–382, 2018. DOI: 10.1177/0959651816678502.

[7] K. López de Calle, S. Ferreiro, A. Arnaiz, and B. Sierra, "Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments", *Computers in Industry*, vol. 112, 2019. DOI: 10.1016/j.compind.2019.07.004.