The Performance of Unsupervised Learning in Predictive Maintenance

Johannes Kriebel[1]

University of Münster, Finance Center Münster, Münster, Germany
johannes.kriebel@wiwi.uni-muenster.de

Extended Abstract

1 Introduction

Predictive maintenance is a smart service that uses predictive tools to determine when to initiate maintenance actions (Carvalho et al. 2019). This work particularly addresses the choice of methods to leverage the benefits of predictive maintenance. Carvalho et al. (2019) and Theissler et al. (2021) recently published literature reviews summarizing commonly applied methods for this domain. Based on these reviews, the most part of extant literature uses supervised learning, while unsupervised learning is used much less often.

In contrast, recent work has highlighted the benefits of using unsupervised learning in predictive maintenance. Brandt et al. (2016) argue that appropriately deployed unsupervised learning could excel in processing vast amounts of sensor data and have benefits in processing data streams. Zschech, Heinrich, Bink & Neufeld (2019) argue that missing or incomplete labels are a typical characteristic of predictive maintenance in real operations, resulting in more demand for unsupervised solutions. Unsupervised learning could potentially also detect infrequent or newly developing patterns (Zipfel et al. 2023).

Based on these results, further research is required to guide the choice between supervised learning and unsupervised learning or combinations of the two. In addition, unsupervised learning does not directly identify failures, but indirectly identifies them via detecting anomalies. While presumably being a small difference, unsupervised models in predictive maintenance need to be designed to identify anomalies that are informative of failures. Identifying more meaningful anomalies could enhance the performance of unsupervised learning in predictive maintenance. The study hence investigates two different research questions:
RQ1: How strong is the performance of unsupervised learning compared to supervised learning in predictive maintenance?
RQ2: How to design unsupervised learning methods that determine meaningful anomalies in predictive maintenance?

2 Conceptual Background: Predictive Maintenance

Maintenance is an important service, enabling companies to remain productive and competitive, and in particular to ensure the functionality and safety of manufacturing (Sang et al. 2020, Zschech, Bernien & Heinrich 2019, Zschech, Heinrich, Bink & Neufeld 2019). The use of connected devices, sensors, and the Internet of Things (IoT), generates vast amounts of data on operating conditions in manufacturing (Gerloff & Cleophas 2017, Zonta et al. 2020, Zschech, Heinrich, Bink & Neufeld 2019, Flath & Stein 2018). Advances in data analytics enable firms to collect and analyze this data and improve maintenance decisions (Sang et al. 2020, Zschech, Heinrich, Bink & Neufeld 2019). Predictive maintenance in this way replaces reactive approaches to maintenance (maintenance in the case of failures) or preventive approaches to maintenance (in regular time intervals), with a proactive approach (Mcloughlin et al. 2022). Benefits of predictive maintenance consist of reduced costs (Ayvaz & Alpay 2021, Kraus & Feuerriegel 2019), but also work place safety and product quality (He et al. 2017, Carnero 2005, Theissler et al. 2021). Given a prolonged lifetime of production resources, it also has implications for sustainability (Ayvaz & Alpay 2021, Kraus & Feuerriegel 2019). A greater reliability can further strengthen customer relationships (Namuduri et al. 2020).

3 Data and Preliminary Results

The data sets for the study were collected by screening the extant literature for publicly available predictive maintenance data. This data particularly consist of the NASA Turbofan Jet Engine Data Set (Kraus & Feuerriegel 2019, Aydin & Guldamlasioglu 2017, Mathew et al. 2017), the NASA Bearing Data (Hong & Zhou 2012), the Hard Drive Failure Data (Byttner et al. 2011), the Autonomous Vehicle Data (Fang et al. 2020), the Safety Pilot Model Deployment Data (Van Wyk et al. 2019), and the Li-Ion Battery Data (Rezvani et al. 2011).

Preliminary results of this extended abstract are further based on the Microsoft Azure Predictive Maintenance data set. The data set contains information on machines with several sensors, maintenance activities, and failures. Based on Carvalho et al. (2019) and Theissler et al. (2021), typical supervised and unsupervised machine learning methods in predictive maintenance were used: logistic regression, random forest, artificial neural network, k-means clustering, and autoencoders. The example further uses an isolation forest that performs favorably in many anomaly detection tasks (Liu et al. 2008). Model performance is evaluated using AUC, Precision, Recall, and F1 metrics on a 70:30 train and test split.

Preliminary results indicate that the methods of supervised learning generally perform slightly stronger compared to the methods of unsupervised learning. Within the supervised methods, the random forest performs particularly strong. A similar picture emerges within the unsupervised learning methods, among which the isolation forest has a particularly strong performance.

The second research question addresses designing unsupervised learning methods that detect anomalies that are particularly informative of failures. This will mainly include using explainable AI methods such as SHAP, but also a more conceptual discussion of what anomalies could appear. In a broader sense, the paper aims to develop a taxonomy of relevant anomalies unsupervised learning methods should address that could be used for model development.

4 Conclusion and Outlook

This work studies the performance of unsupervised and supervised learning methods for anomaly detection in predictive maintenance. While supervised learning has been studied regularly in this domain, unsupervised learning is studied less extensively (Carvalho et al. 2019, Theissler et al. 2021). Yet, recent work has emphasized the importance of unsupervised methods in predictive maintenance (Brandt et al. 2016, Zschech, Heinrich, Bink & Neufeld 2019, Zipfel et al. 2023). The paper investigates the questions how the performance of unsupervised learning compares to supervised learning and how to design unsupervised learning methods to identify informative anomalies. Based on preliminary results, there is a somewhat stronger performance of supervised learning methods. The work collected extensive data to further investigate the comparative performance. Among the unsupervised methods, the isolation forest has a particularly strong performance. The full paper will then use explainable artificial intelligence to determine how methods identify failures in advance and develop a taxonomy of anomalies that are useful in this regard.

References

Aydin, O. & Guldamlasioglu, S. (2017), ‘Using LSTM networks to predict engine condition on large scale data processing framework’, *International Conference on Electrical and Electronic Engineering (ICEEE)* pp. 281–285.

Ayvaz, S. & Alpay, K. (2021), ‘Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time’, *Expert Systems with Applications* **173**, 114598.

Brandt, T., Grawunder, M. & Appelrath, H.-J. (2016), ‘Anomaly detection on data streams for machine condition monitoring’, *IEEE International Conference on Industrial Informatics (INDIN)* pp. 1282–1287.

Byttner, S., Rögnvaldsson, T. & Svensson, M. (2011), ‘Consensus self-organized models for fault detection (COSMO)’, *Engineering applications of artificial intelligence* **24**(5), 833–839.

Carnero, M. C. (2005), ‘Selection of diagnostic techniques and instrumentation in a predictive maintenance program. a case study’, *Decision Support Systems* **38**(4), 539–555.

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Da Francisco, R. P., Basto, J. P. & Alcalá, S. G. S. (2019), 'A systematic literature review of machine learning methods applied to predictive maintenance', *Computers & Industrial Engineering* **137**, 106024.

Fang, Y., Cheng, C., Dong, Z., Min, H. & Zhao, X. (2020), 'A fault diagnosis framework for autonomous vehicles based on hybrid data analysis methods combined with fuzzy pid control', *International Conference on Unmanned Systems (ICUS)* pp. 281–286.

Flath, C. M. & Stein, N. (2018), 'Towards a data science toolbox for industrial analytics applications', *Computers in Industry* **94**, 16–25.

Gerloff, C. & Cleophas, C. (2017), 'Excavating the treasure of IoT data: An architecture to empower rapid data analytics for predictive maintenance of connected vehicles', *International Conference on Information Systems (ICIS)* .

He, Y., Gu, C., Chen, Z. & Han, X. (2017), 'Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis', *International Journal of Production Research* **55**(19), 5841–5862.

Hong, S. & Zhou, Z. (2012), 'Application of gaussian process regression for bearing degradation assessment', *International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM)* pp. 644–648.

Kraus, M. & Feuerriegel, S. (2019), 'Forecasting remaining useful life: Interpretable deep learning approach via variational bayesian inferences', *Decision Support Systems* **125**, 113100.

Liu, F. T., Ting, K. M. & Zhou, Z.-H. (2008), 'Isolation forest', *IEEE International Conference on Data Mining* .

Mathew, J., Luo, M. & Pang, C. K. (2017), 'Regression kernel for prognostics with support vector machines', *IEEE international conference on emerging technologies and factory automation (ETFA)* pp. 1–5.

Mcloughlin, S. J., Puvvala, A., Connolly, N., Maccani, G. & Donnellan, B. (2022), 'Embedding service logic into business model design: The case of predictive maintenance for industry 4.0'.

Namuduri, S., Narayanan, B. N., Davuluru, V. S. P., Burton, L. & Bhansali, S. (2020), 'Review—deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors', *Journal of The Electrochemical Society* **167**(3), 037552.

Rezvani, M., AbuAli, M., Lee, S., Lee, J. & Ni, J. (2011), 'A comparative analysis of techniques for electric vehicle battery prognostics and health management (phm)', *SAE Technical Paper* **191**, 1–9.

Sang, G. M., Xu, L., de Vrieze, P., Bai, Y. & Pan, F. (2020), 'Predictive maintenance in industry 4.0', *International Conference on Information Systems and Technologies* .

Theissler, A., Pérez-Velázquez, J., Kettelgerdes, M. & Elger, G. (2021), 'Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry', *Reliability Engineering & System Safety* **215**, 107864.

Van Wyk, F., Wang, Y., Khojandi, A. & Masoud, N. (2019), 'Real-time sensor anomaly detection and identification in automated vehicles', *IEEE Transactions on Intelligent Transportation Systems* **21**(3), 1264–1276.

Zipfel, J., Verworner, F., Fischer, M., Wieland, U., Kraus, M. & Zschech, P. (2023), 'Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models', *Computers & Industrial Engineering* **177**, 109045.