

Predictive maintenance research at CAISR*

Thorsteinn Rögnvaldsson, Peyman Mashhadi, Yuantao Fan, Mahmoud Rahat,
Reza Khoshkangini, Mohammed Ghaith Altarabichi, Mohamed-Rafik
Bouguelia, Sepideh Pashami, and Sławomir Nowaczyk

Center for Applied Intelligent Systems, Halmstad University, Sweden
{thorsteinn.rognvaldsson, peyman.mashhadi, etc.}@hh.se

Abstract. A brief overview is provided of a few feature representation projects at the Center for Applied Intelligent Systems Research at Halmstad University, with a focus on autonomous knowledge creation and predictive maintenance of machines. We propose different embedded, compressed, disentangled, or transferable representations to enable automatic capture of the encoded knowledge in the features.

Keywords: Machine learning · Feature selection

1 Center for Applied Intelligent Systems Research

The common research vision for Center for Applied Intelligent Systems Research (CAISR)¹ is to do research that can lead to systems that do *autonomous knowledge creation*. The research is done in close collaboration with industrial partners.

A suitable formalism for visualizing how to take steps towards autonomous knowledge creation is the Data, Information, Knowledge, and Wisdom hierarchy [5] – often illustrated with a pyramid as in Fig. 1. The bottom level, *Data*, deals with collecting and representing data. To autonomously select what data to collect, what representations to use, and how to learn general features. The *Information* level relates to questions that begin with “who, what, when, and how many”, creating “events” from the data in the layer below, e.g., rearranging/sorting and aggregating. The *Knowledge* level is about creating rules from the information. This requires combining information from different sources; associating an event from one data source with an event from another data source. The top level, *Wisdom*, relates to the questions “why” or “what will happen?” It is about the ability to project into the future and reason back into the past.

2 Predictive maintenance

Predictive maintenance is about predicting when a system needs to be maintained (e.g., repaired, serviced, old parts replaced, etc.). It builds upon the idea

* Supported by Halmstad University, the Knowledge Foundation, and Vinnova

¹ <https://www.hh.se/caisr>

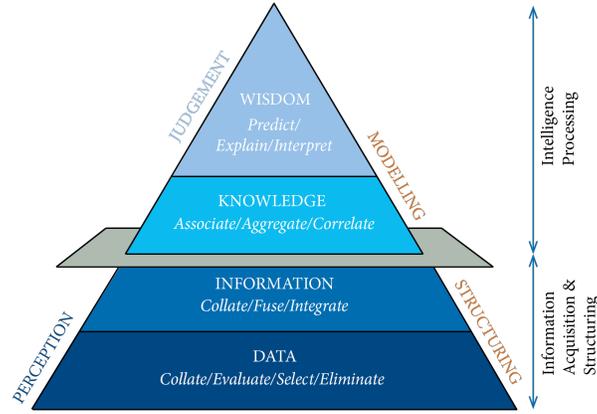


Fig. 1. The Data, Information, Knowledge, and Wisdom (DIKW) hierarchy.

that components or subsystems can be monitored such that it is possible to estimate their health status and predict when they are likely to fail, or to detect when they have failed. Ideally, this removes most unexpected repairs and enables optimization of maintenance operations. Predictive maintenance is an excellent training and testing ground for autonomous knowledge creation. On machines, more and more on-board and off-board data are being collected, documentation of services and repair are becoming better and better, and there is a clear business case for predictive maintenance. This is also visible in the bibliometrics (see Fig. 2): The predictive maintenance field is snowballing, specifically with machine learning tools, and one third of the total number of scientific papers over the last thirty years have been published the last three years.

An important question in predictive maintenance, and in knowledge creation, is to select, construct or learn good features for the data. We provide an overview of five ongoing research projects at CAISR that focus on this.

3 Selecting and generating good features

3.1 Learning embeddings and compressing data

The high dimensional nature of logged vehicle data demands increased storage capacities and processing power on-board vehicles. Besides the available attributes in the LVD are often noisy and redundant in terms of data processing. We have developed a novel way of compressing heavy truck data with the use of artificial neural networks, which achieves 87% compression while improving maintenance prediction by 23% (demonstrated on turbochargers). The approach builds on learning embeddings of multi-dimensional relationships between sensor readings and categorical features.

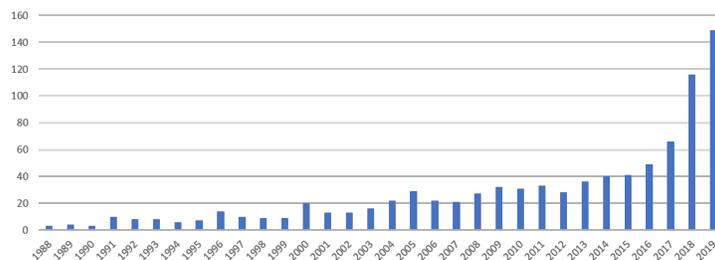


Fig. 2. The number of papers published annually on predictive maintenance, condition monitoring and diagnostics with machine learning techniques.

3.2 Human-computer collaborative feature selection

In several applications when anomalies are detected, human experts have to investigate or verify them one by one. As they investigate, they unwittingly produce a label – true positive (TP) or false positive (FP). We exploit this label feedback to minimize the FP rate and detect more relevant anomalies while also minimizing the expert effort required to investigate them.

Our proposed method iteratively presents the top anomalous instance to a human expert and receives a feedback (positive or negative). Before presenting the next anomaly, the method re-ranks instances such that the top anomalous instances are more similar to TP instances and more dissimilar to the FP instances. This is achieved by learning to score anomalies differently in different regions of the feature space, based on different combinations of features.

Experimental evaluation shows that the method achieves statistically significant improvements on both *detection precision* and *expert effort*, compared to existing state-of-the-art interactive anomaly detection methods.

3.3 Features that build on consensus among models

Many industrial applications operate on fleets of similar equipment. This means that “peers”, i.e. similar equipment with similar usage, can be used for on-line calibration, to estimate the normality of an individual piece of equipment. We use the Consensus Self-Organized Models (COSMO) [1] approach to compute features of streaming on-board data. Such features have been demonstrated to be useful for finding faults and anomalies [3], and also to estimate equipment remaining useful life [2] – including when the operation conditions change.

We use such COSMO features to predict the State of Health (SOH) of Li-Ion drive batteries with real usage data from fleets of buses.

3.4 Parallel Deep Orthogonal Representation Learning

Deep neural networks can be used to learn representations, features, but these features are not “disentangled”, e.g., orthogonal. We have developed a method

where the deep neural network learn disentangled representations in the hidden layers by using a parallel deep architecture with orthogonality constraints.

In practice, this is done by using parallel architectures of deep neural networks, where the units are made orthogonal during training with Gram-Schmidt orthogonalization. This process can be considered as an additional layer of operation after each parallel layer of the model architecture.

3.5 Transferable features

Machine learning models often face a significant transfer challenge in dynamically evolving environments. The conditions under which the model was trained often vary from the testing conditions (the field conditions). Our early work with modeling SoH of Li-Ion drive batteries showed that the deterioration processes of batteries in hybrid buses could vary significantly for different bus configurations and operating conditions, and that many features were not useful (even harmful, leading to negative transfer) to transfer across different settings (e.g., different batteries). Therefore, we have experimented with methods for selecting features that can be transferred from the source domain (training setting) to the target domain (test setting).

We propose to use a Genetic Algorithm (GA) to select invariant features to transfer across multiple source domains D_S . Our work make similar assumption to [4] that if a feature subset is invariant across all source domains, then the same holds across all source and target domains. The *GA* is initiated with a population of individuals encoding feature subsets as chromosomes of binary strings, with 1 indicating inclusion of the feature of the corresponding index. The *GA* evaluates feature subsets according to their performance across all available source domains. Preliminary results show how features can be found that are invariant under, e.g., change of battery generation.

References

1. Byttner, S., Rögnvaldsson, T., Svensson, M.: Consensus self-organized models for fault detection (COSMO). *Engineering Applications of Artificial Intelligence* **24**, 833–839 (2011)
2. Fan, Y., Nowaczyk, S., Rögnvaldsson, T.: Transfer learning for remaining useful life prediction based on consensus self-organizing models. *arXiv preprint arXiv:1909.07053* (2019)
3. Rögnvaldsson, T., Nowaczyk, S., Byttner, S., Prytz, R., Svensson, M.: Self-monitoring for maintenance of vehicle fleets. *Data mining and knowledge discovery* **32**(2), 344–384 (2018)
4. Rojas-Carulla, M., Schölkopf, B., Turner, R., Peters, J.: Invariant models for causal transfer learning. *The Journal of Machine Learning Research* **19**(1), 1309–1342 (2018)
5. Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science* **33**, 163–180 (2007)