CS 680: Project Proposal

John DiMatteo
jdimatteo@uwaterloo.ca

University of Waterloo— October 5, 2018

Abstract

The following is a proposal for a graduate research project for the University of Waterloo Computer Science class "CS 680: Introduction to Machine Learning". The goal is to use machine learning to estimate the operating condition of an asset (called the health score). In particular, the algorithm will be implemented and evaluated on an industrial furnace fan critical to a steel plant. This will allow operators in the steel mill to monitor the asset schedule efficient maintenance tasks, reducing unnecessary maintenance and unexpected failures. A proposed solution, using self-organizing maps, is used to provide asset owners with an indication of the health of a particular asset (known as *condition monitoring*).

Motivation

An often overlooked part of an asset's expenses is maintenance. A popular maintenance strategy is known as Preventative Maintenance (PvM), where maintenance is regularly performed on an asset while it is still in good condition to prevent it from breaking down unexpectedely. Over $ 200 billion is spent on such maintenance every year in the United States and one-third is wasted on improper or unnecessary maintenance [7]. Even worse, no maintenance at all can lead to unexpected failures which in turn can cause serious economic consquences or injury. There exists a significant need to modernize maintenance techniques around the world to ensure safety, reliability, and efficiency.

During the Second World War, a British scientist named Conrad Waddington made a fascinating discovery about the maintenance of aircraft while working for the Royal Air Force (RAF). Previously, aircraft bombers had a notorious problem of breaking down - in fact the ideal serviceabitity in a squadron of bombers was only around 70-75% [8]. What he discovered was that preventative maintenance methods actually increased the rate of unexpected failure. The process of more maintenance leading to more failures became known as the Waddington Effect as a result. By increasing the interval between maintenance cycles and eliminating all maintenance deemed unnecessary, Waddington was able to increase the effective flight hours of the RAF bomber fleet by 60% [8].

After this discovery, asset owners around the world tried to find the optimal time to repair an asset. This led to the invention of Predictive Maintenance (PdM), a philosphy that uses the actual operating condition of assets to optimize operations [7]. For PdM to be effective, the asset's operating condition must be estimated. Estimating the asset's operating condition is the focus of the paper, which is refered to as the *health score*.

Problem Statement

To provide asset owners with a numerical value representing the asset's operating condition or risk of failure.

1 Background

1.1 Why Machine Learning?

One of the most common reasons PdM methods fail is a lack of continuous improvement and a lack of repeatability [3]. Additionally, equipment monitoring is a time consuming process, requires experts to identify failure patterns, and is expensive. Machine learning provides an automated approach that requires minimal asset knowledge, is inexpensive, can be trained on many assets, and can be re-trained as operating conditions change.

Machine learning techniques for predictive maintenance were considered not practical, too complex, or too time consuming. In particular, plant managers did not want to change their existing infrastructure (the software that handles data acquisiton and analyzes it) to adopt the technology. But now Asset Performance Management (APM) software providers are growing and condition based maintenance is at the forefront. As they team up with cloud based data solutions, it becomes easy for asset operators to implement machine learning in their existing data infrastructures via a simple call to the cloud. Manufacturers around the world use APM technology from Bentley Systems, a global leader in APM capabilties according to a recent Gartner report [1]. The proposed solution will be deployed and maintained using APM software from Bentley Systems.

1.2 ArcelorMittal Dofasco

ASK DAVID IF THIS IS OKAY ArcelorMittal Dofasco is a steel company located in Hamilton, Ontario. They use Bentley's APM software and are eager for a machine learning solution to detect the operating conditions of various assets. In particular, they have offered a real data set of several industrial level furnance fans located in the Hamilton plant. Specifically, the data is composed of several smaller data sets, each representing various hours of operation. See the Appendix for full list of variables included.

In a steel making plant, called a steel mill, operations run almost 24/7 except when the mill is shut down once a month for repairs and maintenance. A failure of an asset leading to a shutdown at any other time results in severe costs. If the operating condition of the asset is known, then operators can determine whether or not it should not be repaired during that scheduled downtime, thus avoiding costly unexpected failures and the Waddington Effect.

A reheating furnace is used to raise the internal temperature of steel, so that it can be shaped into a final product. Setting the correct temperature is one of the most essential factors of product quality in the plant. If at any point the blast furnaces fail, the entire line must be shut down to allow operators to repair the furnaces. Operators suggest this could take X amount of time, resulting in approximately $ of lost production **GET A QUOTE FROM DAVID!**.

2 Previous Work

Using machine learning for condition monitoring is not new. There are typically three approaches:

1. Supervised algorithms using sensor data and maintenance data.
2. Unsupervised algorithms using only sensor data.
3. Semi-supervised algorithms using only *healthy* sensor data.

Supervised methods require labeled data. Labeling is the process of associating an output with a set of variable values at a certain point in time. For example, if there is knowledge to when the failures occured and there are sufficient failures, the data can be labeled as healthy or not. Unfortunately, there is just not enough failure data in practice to make this feasible. Instead, failure data can be estimated and artifically generated [10]. But even then the algorithms are limited to a binary outputs.

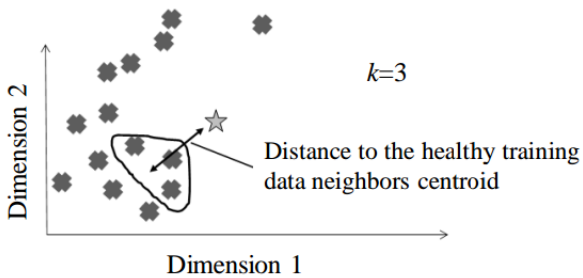
Using k-nearest neighbours (kNN) to estimate asset condition directly is difficult due to noise, and requires domain specific knowledge to choose appropriate variables. One paper improves on kNN methods for detecting the levels of severity for cracks in gears [5]. The disadvantage in these approaches is that it requires significant data in a variety of conditions, and it uses classification to identify a severity level instead of a numerical value.

Researchers in [9] train an autoencoder on healthy imaging data to get a feature representation in a smaller number of dimensions. Utilizing the Support Vector Machine (SVM), they learn a decision boundary to identify anomalies. This approach is interesting for high dimensional data but focuses on point anomalies, where as we are interested in projecting asset health over time.

A self-organizing map is a type of neural network that is often used as a dimensionality reduction technique as it produces a low dimensional representation of the training samples [6]. Training a SOM on healthy asset data, researchers were able to output a health indication in an interpretable way but still lacked a quantifiable measure [2]. Using a measure such as the minimum quantization error, Huang et al. were able to use it as the health indicator for ball bearings [4].

$$Q = \min_{k} \|D - B_k\| \tag{1}$$

Where Q is the minimum quantization error, D is a test set observation, and B_k is the weight vector of the k^{th} closest neuron of the SOM. Using quantization error can be improved as it is sensitive to noise. Researchers Tian et al. use a SOM and a k-nearest neighbours algorithm 2 in combination with the Euclidean distance between the test data and the healthy data to develop a more robust health score.



3 Implementation

The proposed solution is to use a semi-supervised SOM to learn the operating space of the asset during healthy operation. To get a meaningful performance indicator, the distance from a cluster to a datapoint will be used as the health score. The assumption is that a machine that is trending towards failure will deviate more and more from the healthy operating conditions, thus creating a higher health score. The higher the health score, the lower the operating condition of the asset. As discussed, this has been done before but the use of these algorithms on real and significant data is challenging. To improve upon these methods, a decision boundary will be learned as a final step via the use of SVM as seen in previous unrelated work [9].

4 Evaluation

Evaluation will be performed against each of these three scenarios.

1. Run-to-Failure: how will the proposed solution compare if no maintenance is performed?
2. Preventative: how will the proposed solution compare if PvM is performed (the current process of the plant)?
3. Other Predictive Methods: how will the proposed solution compare if another PdM solution is performed?

Metrcs for evaluation: downtime (or uptime), cost, and number of failures prevented. Finally, the algorithm will also be scrunitized by a subject matter expert (SME) in terms of usability and interpretability.

References

[1] Nicole Foust and Kristian Steenstrup. *Market Guide for Asset Performance Management Software*. Jun 2018.

[2] O. Geramifard, J. . Xu, C. K. Pang, J. H. Zhou, and X. Li. Data-driven approaches in health condition monitoring — a comparative study. In *IEEE ICCA 2010*, pages 1618–1622, June 2010.

[3] Douglas Hart. Predictive Maintenance - Avoiding the Ten Most Common Pitfalls. Technical report, Emerson, 02 2017.

[4] Runqing Huang, Lifeng Xi, Xinglin Li, C Richard Liu, Hai Qiu, and Jay Lee. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. 21:193–207, 01 2007.

[5] Yaguo Lei and Ming J Zuo. Gear crack level identification based on weighted k nearest neighbor classification algorithm. *Mechanical Systems and Signal Processing*, 23(5):1535–1547, 2009.

[6] Self-Organizing Map and Teuvo Kohonen. Self-organizing map. *Proceedings of the IEEE*, 78:1464–1480, 1990.

[7] R.K. Mobley. *An Introduction to Predictive Maintenance*. Plant Engineering. Elsevier Science, 2002.

[8] Philip M. Morse. OR in World War 2. Operational Research Against the U Boat. *Science*, 184(4144):1364–1365, 1974.

[9] Philipp Seeböck, Sebastian Waldstein, Sophie Klimscha, Bianca Gerendas, René Donner, Thomas Schlegl, Ursula Schmidt-Erfurth, and Georg Langs. Identifying and categorizing anomalies in retinal imaging data. 12 2016.

[10] Vasilis A. Sotiris, Peter W. Tse, and Michael G. Pecht. Anomaly detection through a bayesian support vector machine. *IEEE Transactions on Reliability*, 59:277–286, 2010.

Appendix

Variable Name	Description
F1S F1SFIBV Overall (g RMS pk)	Furnace 1 south fan inboard bearing vibration
F1S F1SFOBV Overall (g RMS pk)	Furnace 1 south fan outboard bearing vibration
F1S F1SSMIBV Overall (g RMS pk)	Furnace 1 south fan motor inboard bearing vibration
F1S F1SMOBV Overall (g RMS pk)	Furnace 1 south motor outboard bearing vibration
F1S F1NFIBV Overall (g RMS pk)	Furnace 1 north fan inboard bearing vibration
F1S F1NFOBV Overall (g RMS pk)	Furnace 1 north fan outboard bearing vibration
F1S F1NMIBV Overall (g RMS pk)	Furnace 1 north fan motor inboard bearing vibration
F1S F1NMOBV Overall (g RMS pk)	Furnace 1 north fan motor outboard bearing vibration
F1S North Fan Impellor side bearing temp	Furnace 1 north fan outboard bearing temp
F1S North Fan Motor side bearing temp	Furnace 1 north fan inboard bearing temp
F1S South Fan Impellor side bearing temp	Furnace 1 south fan outboard bearing temp
F1S South Fan Motor side bearing temp	Furnace 1 south fan inboard bearing temp

Table 1: Variable names and descriptions provided by subject matter expert at ArclorMittal Dofasco plant.