Machine learning-enabled modeling approach for predictive maintenance decision-making support

Yang, Chunsheng; Yang, Yubin; Yang, Xiaohui; Cheng, Qiangqiang

e-proceedings of ESREL 2020 PSAM 15, 2020-11-05

Machine Learning-enabled Modeling Approach for Predictive Maintenance Decision-making Support

Chunsheng Yang[1], Yubin Yang[2], Xiaohui Yang[3], and Qiangqiang Cheng[4]

[1]*National Research Council Canada, Ottawa, Ontario, Canada*
E-mail: chunsheng.yang@nrc-cnrc.gc.ca
[2]*The State Key Laboratory for Novel Software Technology, Nanjing University, China*
E-mail: yangyubin@nju.edu.cn
[3]*The school of Information and Computer, Nanchang University, China*
E-mail: xiaohuayang@ncu.edu.cn
[4]*The School of Measuring and Optical Engineering, Nanchang Hangkong University, China*
E-mail: chengqq@nchu.edu.cn

Abstract

Modeling system behavior plays a vital role in controlling systems and in monitoring systems health. Recently machine learning-enabled modeling technology has become a powerful technique and tool for developing models for explaining, predicting, and describing system behaviors. In particular, the machine learning-enabled predictive modeling methods have been widely applied to develop the data-driven models from the "big" data in different applications such as system prognostics, system control, and system health management. Over last decade, we have worked on a research program, machine learning-enabled modeling technologies, focusing on the application of machine learning to real-world problems such as prognostics, predictive maintenance, and fault diagnostics. This paper presents the developed machine learning-enabled modeling approaches for predictive maintenance, addressing the challenges such as machine learning algorithm selection, model evaluation, model deployment and decision-making support. The developed models can predict the failure before it occurs and estimate the remaining useful life for predictive maintenance decision-making support. After briefing introduction of the machine learning-enabled frameworks/methodologies, the paper will present a show case of real-world applications, train predictive maintenance decision making support, to demonstrate the feasibility and usefulness of the proposed approaches for predictive maintenance decision making support.

Keywords—machine learning, predictive modeling, lifetime learning, predictive maintenance, decision-making support.

Introduction

To increase the availability and reliability of complex systems such as aircraft and trains and reduce the maintenance costs, we have been working on machine learning-based predictive modelling technologies for predictive maintenance (PM) of complex systems in order to explore a solution for guaranteeing continuous and reliable operations of complex systems. Over past decades we have developed a machine learning-based predictive modelling technology (Yang et al, 2018, 2017, 2014, 2005; Letourneau et al, 2008, 2007, 2005) for developing predictive models for PM. With our modelling technologies, predictive models for a wide range of equipment types can be built from historic operational and maintenance data. The predictive modelling technologies have resulted the advanced PM systems for several real-world applications: aircraft gas turbine engine prognostics (Letourneau et al, 1999), train wheel failure prediction, building operation and maintenance (Yang et al, 2019; Bunay et al, 2019, 2018; Yang et al, 2018, 2017), energy peak prediction for smart buildings, power transformer PM for smart grid (Li et al, 2018) and other applications.

Machine learning-enabled predictive maintenance is to apply machine learning techniques to "big" data analysis for developing predictive models which are able to forecast the status of complex systems after deploying it into the real time monitoring systems. In general such predictive models are developed from the historic data by applying machine algorithms. The main techniques consists of classification and regression. The classification algorithms are used to build classifiers which can be used as the predictive models. Once the models are deployed onto the monitoring system they are able to identify the patterns of any failures and classify the given instances into positive or negative status; other technique is the regression which is used to develop predictive models for trending the performance or estimate the remaining useful life or time to failure. With the outcome of these two kinds of the predictive models, the predictive maintenance (Griall et al, 2002) or proactive operations can be performed in real world applications through decision-making support function or systems. As a consequence, the prediction outcomes can provide necessary information for supporting the decision-making on proactive maintenance, including when to perform the scheduled action for fixing the problems along with confidence explanation on the decision-making support.

This paper presents the developed framework for predictive maintenance by applying machine learning techniques. Following introduction, the paper introduces the machine learning-enabled framework; Section III presents a show case of predictive maintenance to demonstrate the usefulness of the developed techniques; Section IV concludes the paper.

Machine Learning-enabled Framework

Over the last decade we have developed a machine learning-enabled methodology/framework for predictive maintenance decision-making support. The framework consists of three procedures: predictive modelling, model deployment, and decision-making for predictive maintenance, shown as Figure 1. We describe these procedures in this section.

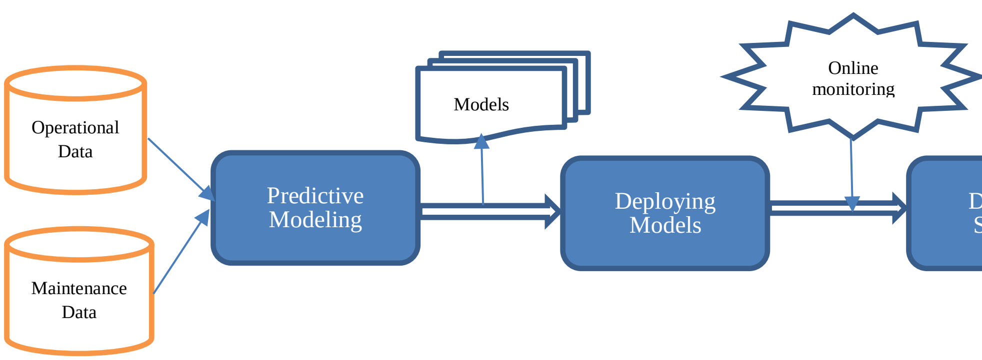

Figure 1, The framework of predictive maintenance

1. Predictive modelling

Predictive modeling in the developed machine learning-enabled framework is a core, and its main task is to develop predictive models for decision-making support by applying machine learning techniques. The models are built from the historic data (operational data and maintenance data) using different machine learning algorithms. To build high-performance models, we developed a predictive modeling methodologies. With the methodologies we can develop two types of the predictive models: model for failure/fault prediction and models for estimating the remaining useful life (or time to failure). For fault predictive model, we casted it as a binary classification problem; for estimation of remaining useful life we casted it as a regression problem in predictive modeling.

A. Classification-based predictive modelling

To build a classifier for failure predictive, we develop a 4-steps method based using machine learning algorithms and data mining techniques. The four steps are data preprocessing, data transformation, modeling and model evaluation.

a) Data pre-processing

To conduct the machine learning-enabled predictive modelling, the first task is to get the relevant data from "big" historic data. The relevant dataset is "gold" data for modelling. There is a thumb rule, called "garbage in garbage out", which means that without the right/relevant data, the models trained with the irrelevant data will mislead the prediction. In terms of the requirements of PM, we are only interested in the data related to failures. For each failure, we create a time series which starts from the onset of a failure and rectorates back certain time window depending on applications. For example in train wheel prognostic, we select data back to 200 WILD (Wheel Impact Load Detector) sites (Yang and Letourneau 2005) from the onset of a failure. Usually the failure events are collected from maintenance database. By retrieving failure events we can get the relevant data from operational database to create a set of time series corresponding to each failure. The time series will be assigned a unique ID to differ from each failure for data engineering.

b) Data engineering

To apply machine learning to predictive modelling is to generate more features for augmenting the data quality in order to build the high performance models. Usually there are two strategies: creating new features from the domain application by using domain knowledge and extracting/learning new features to enhance the data quality. Accordingly, we rely on constructive induction to create new powerful features that combine the initial ones. For feature engineering, it is possible to perform time-series processing to extract potentially relevant time-series characteristics. For instance, in our previous work, to help characterize short-term trends for each axle, we augment the representation with a new feature that stores the moving average of the five most recent measurements of the given axle. In addition to moving averages, we also apply linear regression and Fast Fourier Transforms. We run these techniques in batch mode to generate new features and then add them to the representation. Finally, we apply feature selection (Hall, 2000; Kira and Rendell, 1992) on the augmented data representation to automatically remove correlated or irrelevant features. Recently the deep learning is also used to learn the features automatically.

After we added new features to enhance the data representation and quality, we have to assign the ground truth for each sample in order to build predictive models by applying supervised learning techniques or algorithms. For classification task, the ground truth is a Label (class value) for each instance in the data set. The best way to do this is to consult the domain expert for label value of each sample. However this is costly and not feasible in practice. Therefore, we developed a time window-based labelling method. The idea behind this method is that the observable states of a physical component should be abnormal before a failure occurs. These abnormal states should be labelled as "positive" (class value 1) and the normal states as "negative" (class value 0).

c) Algorithms selection

To build machine learning-based models, many algorithms are available in the research community. There is no theory or guideline to help us to select the suable algorithms to develop the models. This is an empirical method by running large-sale experiments to build a number of models. For classification task, the available algorithms include Instance-Based learning (IBk), TFIDF classifier, Naïve Bayes, Support Vector Machine (SVM), Decision Trees, Neural Networks and deep networks such CNN and RNN. For regression task, there are also many algorithms available, including classic regression, random forest decision tree, SVM, and deep nets. The regression algorithms are mainly used to develop predictive models for estimate the remaining useful life time or time to failure (Grotteke and Trivedi, 2005; Promeller, 2006). For instance we have developed on-demand regression and two stage regression models for estimating time to failure (TTF). However, in practice, these algorithms have to be customized to configure their parameters carefully to get better performance models. To customize an algorithm is fully a data-depending task.

d) Model evaluation

As we described above, there are many models built from the different algorithms and different subset features during the modelling process. Therefore we need an evaluation method to help selecting the suitable mode for deployment in solving real-world problems. There are many criteria or metrics for model evaluation in machine learning community. For classification tasks, these evaluation methods or matrixes are called generic methods, including Accuracy (Error rate), recall, precision, ROC (Receiver Operating Characteristics) curve, AUC (Area Under the ROC Curve), ROCCH (ROC Convex Hull), the DEA (Data Envelopment Analysis), etc. For regression task, the evaluation metrics, including the mean absolute error (MAE), the sum of squared error (SSE), the mean squared error (MSE), and the root mean squared error (RMSE), and the mean absolute percentage error (MAPE). However, these generic methods do not take the specificities or settings of real-world applications into consideration, in particularly, for prognostic applications. The evaluation of the PM models needs to take domain specificities into account. Such specificities cover two aspects: capability of failure prediction and time to failure (TTF) estimation. From the point of view of TTF, it is desirable that a predictive model can generate alerts in a "targeted" time window prior to a failure. A model that predicts a failure too early leads to non-optimal component use. On the other hand, if the failure prediction is too close to the actual failure then it becomes difficult to schedule an effective maintenance. For this purpose, we have developed two evaluation methods for PM model evaluation. The first method is a score-based approach which has been developed by incorporating the problem coverage and time-to-failure into classifier evaluation. This method computes the score based on the testing dataset. The second method is cost-based method which computes the return of investment. When the accurate cost information is available, we can use cost-saving metric to estimate the business value for end users. By using this metric, four kinds of cost information are requested: the cost of a false alert (an inspection without component replacement), a pro rata cost for early replacement, the cost for fixing a faulty component, and the cost of an undetected failure (i.e., a functional failure during operation without any prior prediction from the prognostic model). The first three costs are generally easy to obtain while the last one is difficult to approximate accurately. This is because failures during operations may incur various other costs that are themselves difficult to estimate.

B. Regression-based predictive modeling

Another type of predictive models is the regression-based models for estimating the remaining useful life or time to failure (TTF). To build such predictive models, we developed an On-demand regression (Yang et al. 2017) modeling methods, as shown in Figure 2. This name stresses the fact that regression is only used after a need has been identified. First, the classification-based predictive model analyzes the sensor data from the system to determine if there is a potential failure for monitoring components. When such a risk has been confirmed, then regression is applied. In other words, a potential failure is defined as a positive prediction from a classification-based prognostic model (classifier). To help mitigate the potential negative effects of the issues mentioned in the previous section, several regression models are built and only the most relevant one, which is selected based on evaluation performance of each individual model based on MSE metrics, is applied at any given point in time. As explained below, data clustering plays a key role while constructing the regression models and a simple classifier is used to select the most adequate regression model at run-time. The final step of the proposed approach combines predictions from the classification and regression models. To perform the on-demand regression task for estimating the remaining useful life, two types of the data-driven models, either classifiers or baseline regression models, have to be developed using modeling approach described below from historic operational and maintenance data.

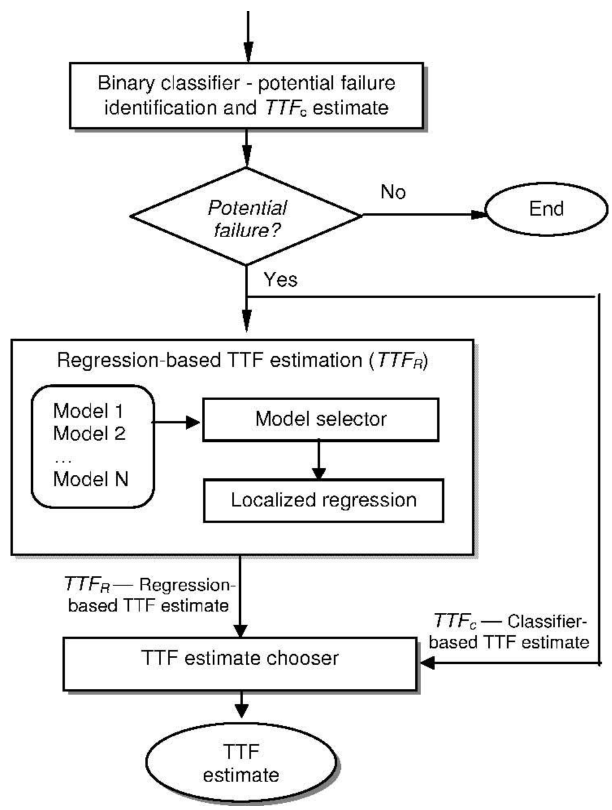

Figure 2, Overview of the on-demand regression method

Two TTF estimates are produced from each positive prediction: one from the classifier (TTF_C) and one from a local regression model (TTF_R). We now need to decide how to combine them into a single TTF estimate. Our approach is straight forward; it returns TTF_R if $TTF_R < TTF_C$. This is to avoid potentially significant errors that could come from an extrapolation of a regression model. TTF_C corresponds to the value used to limit the range of the output attribute while learning the regression models

2. Deploying the built models

Once we selected the best or suitable models from a set of the models available from modeling we need to deploy the model into real-world applications. In general, the model deployment is a type of software engineering problem in the implementation of the predictive models. Over the past decades, we have deployed the developed predictive models into many PM applications such as train prognostics, engine PM, power transformer PM, and so on. With the deployed models, we can also conducted field trail for the PM systems. One more important function for model deployment environment is to allow the PM system to update the predictive model real time and online without reprogram the system. To this end, we have developed a deployment environment which is able to automatically configure and replace out-date / out-mode predictive models.

3. Decision-making support for predictive maintenance

The decision-making support is effectively to provide the useful information or solution for maintenance practitioners based on the outcomes from the predictive models. Such information or solution depends on the various applications. For control system, such decision-making support could be the control policy and control parameter adjustment; For diagnostics, such information could be the fault reason and repair action; For building HVAC (heating, ventilation, and air-condition) facility, such information could be the setpoint change policy, failure description ; For system maintenance, such decision support information could be the confidence of failure prediction, the time to failure from onset of the failure, and the recommended action for fixing the failure before it happen. For the recommended action, it related to scheduling, operation, event optimization. In this work, we only propose to provide the information for predictive maintenance based on the real-time monitoring in practice. In the following case study, we demonstrate how to implement the decision-making support based on model prediction.

A Case Study

A. Background and dataset

Wheel failures, which account for half of all train derailments, cost billions of dollars to the global rail industry. They also accelerate rail deterioration, sometimes causing the rail to break prematurely. All major railways suffer thousands of rail breaks every year. These breaks are dangerous and very expensive to repair. Moreover, the risk for wheel failures and rail breaks is increasing as global competitiveness pushes railways to use larger and heavier cars. To minimize rail breaks and help avoid catastrophic events such as derailments, railways are now closely monitoring the performance of wheels and trying to remove them before they start badly affecting the rails. The data comes from "Wheel Impact Load Detectors" (WILD) (Yang et al, 2005) installed at strategic locations on the rail network. These detectors measure the vertical force or impact of each passing wheel. A central system receives the data in real-time and advises the staff when the impact of a given wheel is too high. The action proposed by the system depends on the actual WILD measurement. In particular, whenever a wheel's impact reaches 140kips11 or more, the train is immediately stopped and then allowed to crawl slowly to the nearest siding, where the car with the defective wheel is uncoupled from the train and is left until a work crew comes to replace the wheel. This process allows the railway to avoid operating with problematic wheels but it introduces other problems: it delays the

[1] A kip is a unit of weight equal to 1,000 pounds. A 140kips wheel impact load is a combined static and dynamic force of 140,000 pounds exerted by an individual wheel at the wheel/rail interface when the freight car passes over the WILD site

offending train and potentially many other trains, it disrupts the overall schedule, and it reduces the railway's throughput capacity. Meanwhile the cargo in the car remains undelivered with potentially negative consequences to the customer. To avoid these problems, railways look for a prognostic approach to predict high impact wheel events sufficiently in advance so that remedial actions can be taken before loading the car and sending it for delivery.

For this purpose, we initialized a research project, WILDMiner, targeting the development of data-mining-based models for prognostic of train wheel failures. The objective is to reduce train wheel failures during operation which disrupt operation and could lead to catastrophes such as train derailments. The data used to build the predictive models come from the WILD data acquisition system. This system measures the dynamic impact of each wheel at strategic locations on the rail network. When the measured impact exceeds a pre-determined threshold, the wheels on the corresponding axle are considered faulty. A train with faulty wheels needs to immediately reduce speed and then stop at the nearest siding so that the car with faulty wheels can be decoupled and repaired. A successful predictive model would be able to predict high impacts ahead of time so that problematic wheels are replaced before they disrupt operation. For this study, we used WILD data collected over a period of 17 months from a fleet of 804 large cars with 12 axles each. After data pre-processing, we ended up with a dataset containing 2,409,696 instances grouped in 9906 time-series (one time-series for each axle used in operation during the study). We used 6400 time-series for training (roughly the equivalent of the first 11 months) and kept the remaining 3506 time-series for testing (roughly the equivalent of the last 6 months). Since there are only 129 occurrences of wheel failures in the training dataset, we selected the corresponding 129 time-series out of the initial 6400 time-series in the training dataset. We created a relevant dataset for modeling which contains 214364 instances from the selected 129 time-series.

B. Model development

Using the obtained dataset, we built two types of predictive models following the framework presented above. The first predictive models are classification-based models which able to predict the failure give a required time window along with confidence level. This predictive model can detect 95% failures with 5% false alert rate. The second type of the models are baseline regression model and a mapper as indicated in Figure 2. In this application, we build 4 baseline regression model and one 4-class classifier for mapper which is used to select the baseline regression model which will estimate the FFT for the predicted failure. The details are referred in (Yang, et al, 2017). With this on-demand TTF predictive system, we observed a much better fit between the estimates and the actual TTF values. With an average error of 4.9±5.4 hours and a minimal squared error (MSE) of 228.

C. Model deployment

After we built the predictive models we implemented the models as the reasoning systems and integrated with the client's IT systems to provide the decision making for predictive train maintenance. We used the model deployment tool as described above. This tool allows the PM system to update the predictive model real time and online without reprogram the system and is able to automatically configure and replace the outdate /outmode predictive models. The decision-making support system is a web-based application, which can be accessed at anywhere and anytime. The following is the detail description on the decision-making support for predictive maintenance.

D. Decision-making support for predictive maintenance

Based on the outcome of the predictive models, the failure prediction and TTF estimation, we developed a web-based decision-making support system. In this work, the failure prediction from the classification-based model is "failure" with a confidence level. A TTF prediction is converted as a remaining distance before failure happens. The Figure 3 is the main window of the decision-making system, which displays the list of the failure alert form the operation. Each line represents an alert (a failure prediction), which contains necessary information for decision-making support, including the location, date time, priority, confidence impact, and so on. For more information for each alert, the operator can click the alert to launch the explanation information or recommendation window.

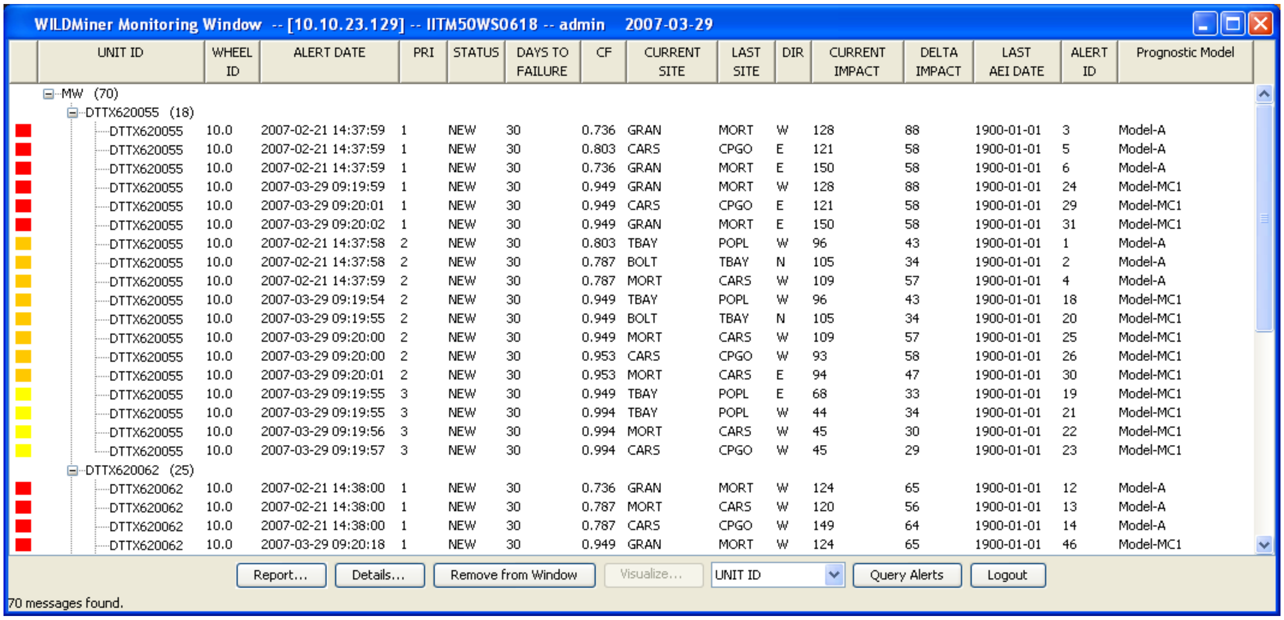

Figure 3, the main window of decision-making support

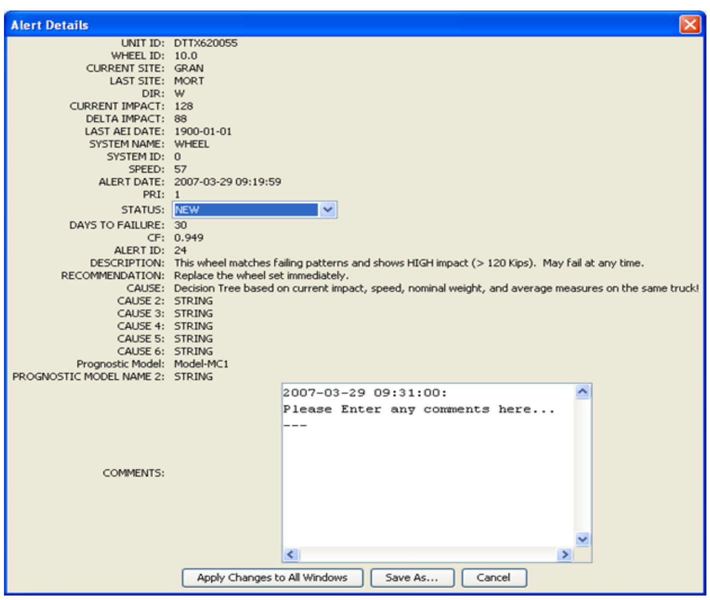

Figure 4, the explanation information for given alert

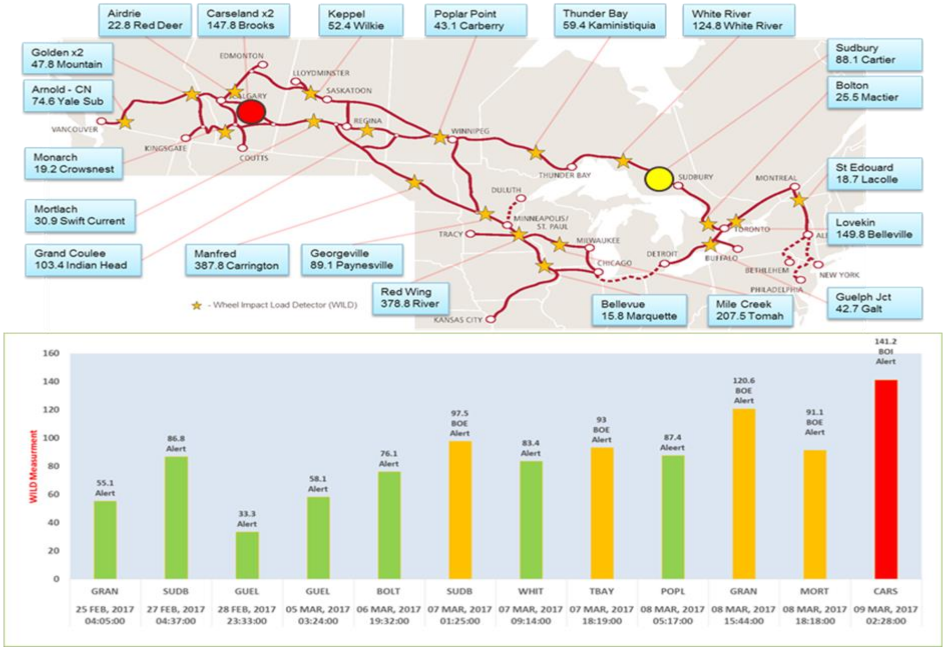

Figure 5, an example with more detailed decision-making support information

Figure 4 shows an explanation information window for a given alert. It provides more detail information from the main window. The details include the reason of a failure, maintenance request/record, and recommended repair action. For example, it indicated that the high impact (> 120 kips) is predicted and may fail any time. The recommendation action is "to replace the wheels immediately".

Figure 5 shows an example of detailed information for decision-making support. The part above in Figure 5 shows the WILD sites location and rail network, the part below shows the ground truth and model prediction for each WILD site. The red bar shows the failure date0time, and yellows are the prediction from the predictive models, the green bars shows the normal operations for a given train. In this case, a train departed on Toronto Station and head for Vancouver. On March 7, 2017, 01:25AM the train passed WILD site, Sudbury the predictive model forecast a potentials failure with estimated remaining distance is 1750 miles (about 29 hours). Based on the real date the failure happened at WILD site, CARS, on March 9, 2017, 02:28AM. Before the operators employ the predictive maintenance support system, they had to set out the train at Calgary station when failure happened which cause a considerable service delay and operation cost. With the help of the predating results from the predictive models, now the operator can schedule the maintenance service before the failure happen with planed service for unloading the goods in 9 hours and largely reduced the service delay and saved the cost as well.

Conclusions

In this paper, we presented a machine learning-enabled framework for predictive maintenance decision-making support. We have developed machine learning-based methodologies for building predictive models from big historic data. We also presented a show case to demonstrate the feasibility and usefulness of the developed framework. The show case is from the real-world application, train predictive maintenance. The field experiments indicated the developed predicted model can provide the accurate failure prediction with relative precision of TTF estimation. These outcomes are useful for providing right information for train predictive maintenance and can reduce largely the operation cost.

Acknowledgement

This research is supported by a new ideation program fund from the National Program Office at the National Research Council Canada. It is also partially supported by National Science Fund Foundation (Grant No.61963026). Thanks to all support and value discussion from the colleagues at the National Research Council and other Research Institutes.

References

Li, A., Yang, X., Dong, H., Xie, Z., and Yang, C., Machine Learning-Based Sensor Data Modelling Methods for Power Transformer PHM, *Sensor* Vol. 18, No. 12, Dec, 2018, pp. 4430-4440

Yang, X., Chen, W. , Li, A., Yang, C., Xie, Z., and Dong, H. BA-PNN-based methods for power transformer fault diagnosis, *The Journal of Advanced Engineering Informatics*, Vol. 39, Jan, 2019, pp. 178-185

Gunay, B., Shen, H. and Yang, C., Text-mining building maintenance work orders for component fault frequency, *Journal of Building Research & Information*, Vol. 47, No. 5, 2019

Gunay, B. Shen, W., Huchuk, B., Yang, C., Bucking, S. and O'Brien, W. Energy and comfort performance benefits of early detection of building sensor and actuator fault, *Journal of Building Services Engineering Research & Technology*, online first, 2018

Gunay, B., Shen, W. and Yang, C. Blackbox modelling of the central heating and cooling plant equipment performance, *Science and Technology for the Built Environment*, 24:4, pp.396-409, 2018

Yang, C., Lou, Q., Liu, J, Yang, Y., and Cheng, Q. Particle Filtering-based Methods for Time to Failure Estimation with a Real-world Prognostic Application, *International Journal of Applied intelligence*, Vol.48 No. 8, 2516-2526, 2018

Yang, C. Shen, W., Chen, Q.,, and Gunay, B. A practical solution for HVAC prognostics: Failure mode and effects analysis in building maintenance, *Journal of Building Engineering* , Vol 15, pp 26-32, 2018

Yang, C. Trudel, E. and Liu, Y. Machine Learning-based Methods for Analysing Grade Crossing Safety, *Cluster Computing: the Journal of Networks, Software Tools and Applications*, Volume 20, pp 1625-1635, 2017

Yang, C., Létourneau, S. Liu, J., Cheng, Q, and Yang, Y. Machine learning-based methods for TTF estimation with application to APU prognostics, *International Journal of Applied intelligence*, Vol. 46, No. 1, pp 227–239, Jan, 2017

Yang, C., Zou, Y., Lai, P., and Jiang, N. Data mining-based methods for fault isolation with validated FMEA model ranking, *International Journal of Applied intelligence*, Vol.43, No.4, 2015, pp.913-923

Yang, C., Létourneau, S., and Guo, H. Developing data-driven models to predict BEMS energy consumption for demand response systems, T*he Proceedings of the 27th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (IEA/AIE-2014)*, Taiwan, June, 2014. Pp.188-197

Yang, C. and Létourneau, S. Learning to predict train wheel failures, in Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2005), Chicago, USA, August 2005, pp. 516–525.

Grall, et al., Continuous time predictive maintenance scheduling for a deteriorating system, *IEEE Transaction on Reliability*, Vol. 51, No 2, pp.141-150, 2002

L´etourneau, S., Yang, C. Drummond, C., Scarlett, E., Valdes, J., and Zaluski, M. A domain independent data mining methodology for prognostics, *in Proceedings of the 59th Meeting of the Society for Machine Failure Prevention Technology*, MFPT '05, Virginia Beach, Virginia, USA, April 2005.

L´etourneau, S., Yang, C., and Liu, Z. Improving preciseness of time to failure predictions: Application to APU starter, *in Proceedings of the 1st International Conference on Prognostics and Health Management*, Denver, Colorado, USA, October 2008.

Létourneau, S., Yang, C., and Liu, Z. On-demand Regression to Improve Preciseness of Time to Failure Prediction, *The Proceedings of the 2007 AAAI Fall Symposium on AI for Prognostics*, Arlington, Virginia, USA, November 9-11, 2007, pp.81-86

Hall, M. Correlation-based feature selection for discrete and numeric class machine learning, in *Proceedings of the 17th International Conference on Machine Learning*, 2000, pp. 359–366.

Kira, K.,and Rendell, L. A practical approach to feature selection, *in Proceedings of the 9th International Conference on Machine Learning*, 1992, pp. 249–256.

Grottke, M., and Trivedi, K. On a method for mending time to failure distributions, *in Proceedings of International Conference on Dependable Systems and Networks, Los Alamitos*, USA, 2005, pp. 560–569

Promellec, B., Riant, I. and Tessier-Lescourret, C. Precise life-time prediction using demarcation energy approximation for distributed activation energy reaction, *Journal of Physics: Condensed Matter*, vol. 18, no. 2006, pp. 2199–2216, 2006