The Application of Random Forest to Predictive Maintenance

Conference Paper · May 2018

Proceedings of the 2018 IISE Annual Conference
K. Barker, D. Berry, C. Rainwater, eds.

The Application of Random Forest to Predictive Maintenance

Rodney Kizito, Phillip Scruggs, Xueping Li and Reid Kress
Industrial and Systems Engineering
University of Tennessee, Knoxville, TN 37996, USA

Michael Devinney and Thomas Berg
Consolidated Nuclear Security, LLC, Oak Ridge, TN 37830, USA

Abstract

Predictive maintenance (PdM) is used to predict when equipment failures may occur and to help provide adequate equipment monitoring and maintenance planning in advance of potential future failure. By using PdM, organizations are able to identify root-causes of failures, minimize the reliability issues of equipment, the frequency of maintenance, and ultimately reduce the potential incurred costs from reactive and excessive preventive maintenance actions. This paper discusses a preliminary study performed to highlight the potential of applying a classification and regression-based random forest (RF) algorithm for predicting the failure of equipment (for the classification approach) and remaining useful-life (for the regression approach). The RF algorithm is trained on a training dataset and then applied to a test dataset to determine the accuracy of the failure prediction.

Keywords
Predictive maintenance, random forest, machine learning

1. Introduction
The prediction of equipment failure can be applied to predictive maintenance (PdM). Accurate failure prediction and remaining useful-life (RUL) prediction of equipments helps to establish adequate maintenance planning. PdM utilizes data sources and a variety of techniques to monitor equipment performance, product quality, and predict eventual equipment failure. By using PdM, organizations are not only able to predict when equipment failures may occur, but also identify the root-causes of the failure and help save on maintenance related costs.

As an example, one established area of PdM focuses on vibrations produced by rotating assets such as motors and pumps. These assets usually drive production or support equipment and often produce stoppages of work for entire facilities during a failure event. A vital component to the operation of rotating assets is the roller bearing. Roller bearing defects can be categorized as point/local defects and distributed defects [1]. These defects can be produced during the initial manufacture of the bearings or through normal use. Local defects include cracks, corrosion pitting, brinelling, and spalls on the rolling surface [1]. Vibration from bearings can generate noise that, if large enough, may degrade the quality of the product line. Severe vibrations may even cause catastrophic failure resulting in an inoperative system and economic loss if the system is unable to be repaired quickly. Monitoring these assets helps provide early detection of faults and allows for the detection of emerging problems before a failure occurs. Machine learning (ML) based predictive algorithms possess the potential to improve predictive models that have been used in the past. One of these algorithms, random forest (RF), and its application towards PdM of equipment, is the focus of this paper. The objective of the preliminary study performed in this paper is to highlight the value of applying RF for the prediction of failure and RUL of equipment. The following section provides background information on the IEEE challenge where the dataset used in this study originated.

2. Background
This study applies a classification and regression-based approach of the RF method to the dataset used in the IEEE 2012 PHM Data Challenge competition. The competition was held by the IEEE Reliability Society and the FEMTO-ST Institute. Each competing team's objective was to provide the best estimate of RUL for ball bearings. The ball bearings were experimentally run to failure under three different operating conditions. Condition 1 used a 4000 N

radial load and operated at 1800 rpm, condition 2 used a 4200 N radial load and operated at 1650 rpm, and condition 3 used a 5000 N radial load and operated at 1500 rpm [2]. Conditions 1 and 2 were applied to seven bearings each while condition 3 was applied to three bearings, resulting in 17 total bearings.

Data acquisition was performed using two accelerometers mounted on the bearings. One accelerometer measured vibrations in the vertical direction while the other measured vibrations in the horizontal direction. Vibration signals were sampled with a frequency of 25.6 kHz and recorded every 10 seconds [2]. Each observation contained 2560 vibration readings. The resulting dataset contained the hour, minute, second, micro-second, horizontal vibration reading, and vertical vibration reading for each observation. Contestants were given 6 bearing run-to-failure datasets to build prognostics models, which were then used to estimate the RUL of the remaining 11 bearings. The following section provides a short review of previous research in PdM and the application of RF to PdM.

3. Literature Review

A short literature review is performed to highlight previous research in the PdM field. The review focuses on PdM in ML through RF. The ML-based PdM field can be modeled in one of two approaches: (1) Classification, which can predict the possibility of failure in the next *n*-number of steps; (2) Regression, which can predict the remaining useful life (RUL) until failure [3]. The independent variable - used to predict the value of the dependent variable - in both modeling approaches is categorical. The difference between the two approaches lies in the dependent variable - the predicted outcome. Dependent variables for Classification are categorical and discrete, while dependent variables for regression are numerical and continuous. The ML method of focus in this paper is RF and is demonstrated from a classification and regression approach. The review includes literature that describes the RF method, along with examples of literature where the method has been applied for PdM.

3.1 Predictive Maintenance

Shin et al. performed a degradation mode and criticality analysis on the product life cycle of a construction equipment vehicle. The analysis was based on data gathered during product usage periods [4]. A condition-based maintenance algorithm was developed to help estimate the RUL of the vehicle's lifting arm. Sensors were used to provide real-time information on the degradation status of the arm. The sensor readings, along with future product usage data, operation data, and working environment data were used to estimate the RUL of the equipment.

Susto et al. presented a multiple classifier ML method for PdM of a generic ion implanter tool [5]. The method improves maintenance management and decision making by introducing classifiers for integral type faults - machine failures due to wear and tear from usage and stress on equipment parts. Each classifier provides information on the health status of the process and of the machine. The methodology uses recorded maintenance data based on how many machine runs occurred leading up to each maintenance cycle. During each machine run, the classifier on each fault returns an F - signifying the iteration was faulty, or an NF - signifying a not-faulty iteration. The objective is to minimize total operation cost, which is composed of two metrics: (1) Frequency of unexpected breaks; (2) Amount of unexploited lifetime on a machine due to failure [5].

3.2 Random Forest

RF is an ensemble learning algorithm based on the 'bagging' method of trees [6]. The trees are independently constructed using samples from the dataset, and a majority vote is taken for the prediction. RF adds randomness to this method of classification trees [6]. This randomness changes how both classification and regression trees are constructed. Each node is split based on the best predictors from a randomly chosen subset of trees. The algorithm draws bootstrap samples from the data, grows a classification or regression tree, and then predicts new data by aggregating predictions from the developed forest of trees. The results include a measure of feature importance (i.e., the value gained by including a certain feature) and a confusion matrix displaying the accurate and inaccurate predictions.

Qin et al. used the RF method to predict the malfunction of wind turbines. The wind turbine data was obtained through the use of the supervisory control and data acquisition (SCADA) approach [7]. An adaptive neuro-fuzzy interface system was applied to a variety of SCADA signals to further mine the data. The RF bagging method was beneficial because of its use of independent learners and its ability to fit data in parallel [7]. The results showed tables of features selected by actual experience and by the RF. Many of the features selected by the two methods were the same, thus showing the quality of RF's feature selection. While using support vector machine (SVM) as a comparative method, RF in general took less computation time and provided a more precise model.

Vantuch et al. utilized RF for predictive analysis on conductor faults by identifying partial discharge activity in fault points [8]. Complex networks were used to extract features for analysis through RF classification. The features were used to measure whether conductor signal contained a fault or not. The positive results obtained showed RF's value for predictive analysis and enticed further research [8].

4. Dataset Development

The original dataset was separated into learning, test, and validation sets. The learning set contained run-to-failure experimental results (vibration observations) for only 6 bearings. The validation and test sets contained identical data thus only the test set is included in the dataset used for this study. The test set contained the remaining 11 bearing run-to-failure experimental results and was combined with the learning set to generate the full dataset of vibration observations. The following subsections explain the actions taken to develop the dataset.

4.1 Attribute Creation

Attributes are created that provide information about the data and help with the eventual feature selection in the RF prediction process. Applying RF allows for the identification of the most important features to the prediction model. These are the features that provide the most information about the occurrence of failures in the equipment. Selecting only the most important features to use for prediction reduces the computation time, increases prediction model's accuracy, and helps avoid over-fitting the model to non-important features. The attributes created in this study are based on the raw horizontal and vertical sensor observation readings. In developing the dataset, the following occurs: (1) Responses are converted to binary outputs where 1 signifies failure and 0 otherwise; (2) The data is balanced and binned based on time (hour-minute-second: HMS) and then based on sampling frequency (SF); (3) Data attributes are created using time domain statistics.

4.2 Output Conversion

Per the IEEE challenge, failure is described as an absolute value of the accelerometer reading (horizontal and vertical) that is greater than 20.00g [2]. The same failure threshold is observed for this study. Both the horizontal and vertical readings are converted to binary outputs. With binning applied, if a value of 20.00 or higher is observed within the horizontal or vertical reading of a bin, then that bin results in a failure (i.e., 1 as the its output). If no value of 20.00 or higher is observed within the horizontal or vertical reading of a bin, then that bin results in a non-failure or 0 as its output.

4.3 Balancing and Binning

Once the output conversion is achieved, it is clear that the dataset used in this study is imbalanced. An imbalanced dataset is one which posses a large difference between the observations of different classes. The classes of focus are the two outputs. The dataset contains a large amount of 'not failed' or 0 observations, compared to the 'failed' or 1 observations. Imbalanced dataset can lead to problematic results because typical ML methods utilize a global search and do not consider the amount difference between classes [9].This imbalance causes the developed ML model to heavily favor the majority class and neglect the minority. To address this problem, a random oversampling technique is applied. Oversampling creates new observations for the minority class in an attempt to re-proportion the dataset to a more balanced one. Balancing is performed pre-binning and again on the binned datasets.

Due to the large size of the dataset (63,715,840 observations) used in this study, binning the data into equal intervals reduces the complexity and computation time of the model. Binning also helps with the discretization process of the continuous dataset, enabling the dataset for numerical evaluation [10]. The first method of binning divides the dataset by the hour-minute-seconds (HMS), grouping all microsecond observations that fall into a specific hour-minute-second frame together. The second method of binning follows the equal width interval binning method [10], which divides the dataset based on the sampling frequency (SF). The given frequency from the competition is 25.6 kHz over a 1/10 s interval, for a total of 2560 observations per interval. For RUL computation using the SF binning method, the bin total (2560) is cut in half (1280) to allow for adequate binning of the smaller datasets.

4.4 Time Domain Statistics

Since the vibration readings are time-dependent, time domain statistics are used to create data attributes. Applying time domain statistics helps remove possible trends within the data. Removing trends, which can be caused by the changes in operating conditions of the bearings and the degradation of the bearings over the experimentation time-frame, makes the data non-stationary and reduces complexity within the model. The statistics are computed for the horizontal and vertical readings separately and combined; the statistics are described as follows [11]:

$$Mean : x_m = \frac{\sum_{n=1}^{N} x(n)}{N} \tag{1}$$

$$Standard\ Deviation : x_s = \sqrt{\frac{\sum_{n=1}^{N} [x(n) - x_m]^2}{N-1}} \tag{2}$$

$$Range : x_r = \| max[x(n)]| - |min[x(n)]| \tag{3}$$

$$Peak : x_p = max|x(n)| \tag{4}$$

$$Root\ Amplitude : x_{ra} = \left[\frac{\sum_{n=1}^{N} \sqrt{|x(n)|}}{N}\right]^2 \tag{5}$$

$$Root\ Mean\ Square : x_{rms} = \sqrt{\frac{\sum_{n=1}^{N} x(n)^2}{N}} \tag{6}$$

$$Skewness = \frac{\sum_{n=1}^{N} [x(n) - x_m]^3}{(N-1)(x_s)^3} \tag{7}$$

$$Kurtosis = \frac{\sum_{n=1}^{N} [x(n) - x_m]^4}{(N-1)(x_s)^4} \tag{8}$$

$$Crest = \frac{x_p}{x_{rms}} \tag{9}$$

$$Margin = \frac{x_p}{x_{ra}} \tag{10}$$

$$Shape = \frac{x_{rms}}{x_m} \tag{11}$$

$$Impulse\ Factor = \frac{x_p}{x_m} \tag{12}$$

where n represents the vibration readings and N is the total number of readings. The following section details the RF methods used to perform the analysis in this study.

5. RF Methodology

The RF algorithm is chosen over other commonly used classification and regression algorithms such as SVM and linear regression (LR). RF can be applied to large and uncleaned datasets, which the dataset in this study is, without increasing computational complexity. Such datasets slow the training speed of the SVM algorithm. SVM is more applicable to small-to-medium sized datasets that are outlier free [12]. The LR algorithm requires a linear relationship between the input and output/response variables, a relationship that RF does not require, which does not exist for the inputs (attributes) and outputs of the dataset used in this study. Thus, RF is preferred over LR. RF's ability to easily interpret features and their importance also make it the desired algorithm for this study.

The algorithm uses the binary outputs to classify the observations as failed or not failed. The regression estimate of the output is used for computing RUL. Bootstrap aggregating (bagging) is used to reduce variance and tree correlation, while leaving prediction bias unaffected. Each tree model is grown on an independent bootstrap sample, which is randomly chosen with replacement, from the data. Bootstrap samples create new training sets of predictors by randomly sampling the given (original) set $M' \leq M$. The predictors not used in the bootstrap sampling are referred to as 'out-of-bag' or OOB. Voting (classification) or averaging (regression) the predictions of these OOB trees provides the RF predictor [13]. The OOB error rate is the error rate of the RF predictor and the confusion matrix is obtained from the RF predictor [13]. As more tree models are added to the ensemble, the OOB error provides an estimate of the prediction test error. The RF algorithm, as described by Kuhn and Johnson, is presented in Algorithm 1 [14].

Algorithm 1 is general for the classification and regression approach. Classification uses trees and predictors to provide a measure of how much the votes for one class exceed those for other classes within the output/response [15]. This

Algorithm 1 Random Forest

Select the number of models (M) to build
for $i = 1$ to M, **do**
  Generate a bootstrap sample of the original data
  Train a tree model Y_i on this sample
  for *each split* **do**
    Randomly select k of the original predictors
    Select the best predictors among the k predictors and partition the data
  end
  Use typical tree model stopping criteria to determine when a tree is complete (but do not prune)
end

Table 1: RF Classification Results

Bearing	Row Name	HMS Binning		SF Binning	
C1B1	Accuracy	1.0		1.0	
	Not-Failed	701	0	1401	0
	Failed	0	677	0	1372
C1B2	Accuracy	1.0		1.0	
	Not-Failed	204	0	430	0
	Failed	0	200	0	408
C1B3	Accuracy	1.0		1.0	
	Not-Failed	504	0	1028	0
	Failed	0	487	0	1053
C2B3	Accuracy	1.0		1.0	
	Not-Failed	485	0	955	0
	Failed	0	467	0	964

Table 2: RF RUL Regression Results

Bearing	Row Name	HMS Binning	SF Binning
C1B1	Dataset Size	2797	8483
	OOB R-2 Score Estimate	0.97150	0.93127
	Test Data R-2 Score	0.97820	0.92027
C1B2	Dataset Size	296	1093
	OOB R-2 Score Estimate	0.60697	0.42304
	Test Data R-2 Score	0.61276	0.39885
C1B3	Dataset Size	2260	6397
	OOB R-2 Score Estimate	0.94433	0.89754
	Test Data R-2 Score	0.94812	0.90095
C2B3	Dataset Size	1950	5472
	OOB R-2 Score Estimate	0.85363	0.76184
	Test Data R-2 Score	0.84519	0.77089

allows for the prediction of a single class (the class with the most votes). Regression provides a prediction by averaging the overall tree predictions [15]. Both approaches are performed with 100 tree models (estimators) in this study. The more estimators used the longer the computation time but the more reliable the results. The two approaches differ in standards of the RF tuning parameter m_{try}. m_{try} is the number of variables available at each split and is typically set at 1/3 of k for regression and$\sqrt{k}$ for classification, where k is the number of randomly selected predictors [15].

6. Results

Only bearings with more than 100 failed raw data observations are chosen for analysis. These are bearings 1, 2 and 3 under condition 1, and bearing 3 under condition 2. These condition-bearing combinations are referred to as C1B1, C1B2, C1B3 and C2B3 respectively. Table 1 displays the accuracy and confusion matrix results for the RF classification and compares the HMS and SF binning methods for the 4 condition-bearing combinations. The accuracy scores explains the percentage of correct predictions made by the RF while the confusion matrix displays the number of correct predictions to incorrect predictions. For example, the RF correctly predicted 677 of the 677 possible failed observations from the C1B1 dataset, and incorrectly predicted 0 of the 677 possible failed observations. The results in table 1 show that the RF method performed classification extremely well, with an accuracy of 1.0 for each bearing.

Table 2 displays the OOB R-squared score estimate, the test data R-squared score and the dataset size (dataset sizes are post-balancing and binning) for bearings' RUL analysis; R-squared describes the goodness-of-fit of the regression. The results show the RF regression performed better on bearings with larger datasets (C1B1 and C1B3) than smaller dataset bearings (C1B2). This is consistent across both binning methods. For RUL computation, the expectation of the HMS-binned regression method was to provide an estimate of remaining operational time until the first failure of each bearing. RUL was properly computed for each of the bearings, with the remaining operational time reaching 0.0 at the hour-minute-second-microsecond of the first failed observation for each bearing. The RUL computation expectation was the same for the SF binning method, and was achieved as the remaining operational time reached 0.0 at the bin where the first failed response is witnessed.

7. Conclusions

This paper presents a preliminary study performed to assess the potential of utilizing RF for PdM. Classification was used for failure prediction and regression was used for RUL prediction. Although the classification approach resulted in perfect prediction accuracy for the dataset used in this preliminary study, such perfect results are not expected when the approach is applied to other equipment failure datasets. However, the performance of the classification in this study shows the potential of using RF for PdM and is further supported by the consistency in prediction accuracy results across the binning methods.

The HMS and SF binning regressions prove the necessity of using a sufficiently sized dataset due to the greater prediction accuracy for bearings with larger datasets. Since the SF binning method is not based on the actual time (hour-minute-second-microsecond) of the observations, it does not perform as well as the HMS method when computing RUL. The R-squared score estimate for each bearing is relatively identical to the test data R-squared score, which suggests a quality prediction performance by the RF regression regardless of dataset size or binning method.

References

[1] Lee, J., Kao, H., and Yang, S., 2014, "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment," Procedia CIRP, 16, 3-8.

[2] Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N., Varnier, C., 2012, PRONOSTIA: An Experimental Platform for Bearings Accelerated Life Test. IEEE International Conference on Prognostics and Health Management, Denver, CO, USA.

[3] Perera, S., and Alwin, R., 2017, "Machine Learning Techniques for Predictive Maintenance," InfoQ.

[4] Shin, J., Jun, H., Catteneo, C., Kiritsis, D., 2015, "Degradation mode and criticality analysis based on product usage data," Int J Adv Manuf Technol 78:1727-1742

[5] Susto, G.A., Schirru, A., Paampuri, S., McLoone, S., Beghi, A., 2015, "Machine learning for predictive maintenance: A multiple classifier approach," Industrial Informatics IEEE transactions, vol. 11, no. 3, pp. 812-820.

[6] Liaw, A. and Wiener, M., 2002, "Classification and Regression by Random Forest," R News, 2, 18-22.

[7] Qin, S., Wang, K., Ma, X., Wang, W., and Li, M., 2017, "Chapter 9: STEP Standard in Design and Manufacturing Ensemble Learning-Based Wind Turbine Fault Prediction Method with Adaptive Feature Selection," Communications in Computer and Information Science Data Science, 572-582.

[8] Vantuch, T., Gaura, J., Misak, S., and Zelinka, I., 2016, "A Complex Network Based Classification of Covered Conductors Faults Detection," Advances in Intelligent Systems and Computing Intelligent Data Analysis and Applications, 278-286.

[9] Rio, S. D., LÃşpez, V., Benitez, J. M., and Herrera, F., 2014, "On the use of Map Reduce for imbalanced big data using Random Forest," Information Sciences, 285, 112-137.

[10] Dougherty, J., Kohavi, R., and Sahami, M., 1995, "Supervised and Unsupervised Discretization of Continuous Features," Machine Learning Proceedings 1995, 194-202.

[11] Mallesh, D., 2017, "Developing Leading and Lagging Indicators to Enhance Equipment Reliability in a Lean System," M.S. thesis, The University of Tennessee.

[12] Smola, A. J., and Scholkopf, B., 2004, A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

[13] Cutler, A., 2010, "Random Forests for Regression and Classification," Lecture, Utah State University.

[14] Kuhn, M., and Johnson, K., 2013, Applied Predictive Modeling. Springer, 200.

[15] Random Forests. Retrieved December 21, 2017, from http://www.statsoft.com/Textbook/Random-Forest.