A Jet Engine Prognostic and Diagnostic System Based on Bayesian Classifier

M.Saeidi, M. Soufian*, A. Elkurdi and S. Nefti-Meziani

School of Computing, Science and Engineering, University of Salford
Manchester, United Kingdom

*magid@ieee.org

***Abstract*—In this work, a predictive maintenance system is discussed as a modern solution for reducing downtimes in complex systems such as airplanes' jet engines. The developed predictive maintenance system is based on prognostic and predictive algorithms which will be constructed by using machine learning techniques. Bayesian theorem is specially studied and employed for this purpose in this paper. The design and implementation of a Naïve Bayesian classifier will be presented to demonstrate and challenge the practicality of the method. A turbofan jet engine health check system was chosen as a complex and live industrial testbed example. We also demonstrate that the system in question has a high entropy and despite this, the Bayesian approach is sufficient enough to eliminate the critical errors as well as maintain a satisfactory overall accuracy.**

Keywords—machine learning, naïve Bayesian, predictive maintenance

I. INTRODUCTION

The advances in engineering and technology has continuously led to automated systems, which are faster, more reliable and efficient [1]. Despite the profound breakthroughs, industrial systems are still subject to unplanned downtime of the machineries and their subsystems. Consequently, these unplanned downtimes have a direct opposing impact on the reliability and profitability of the business [2]. Needless to say, it also can cause catastrophic events, such as loss of lives or huge financial costs. According to a study [3], machinery downtime is costing Britain's manufacturers more than £180 billion every year. To lessen these consequences, an effective maintenance plan is needed to be arranged within the business. Usually maintenance involves functional checks, repair a piece of equipment and so on, and unlike machinery improvement, maintenance has traditionally been considered an expense to minimize [4].

The traditional maintenance action plan approaches can be classified as a) corrective maintenance that performed on the event of a failure, and b) scheduled or periodic maintenance that performed at a regular base and well ahead of any sign of fault. Scheduled maintenance involves checking, repairing or replacing all the elements that are subject to wear without reaching their end of life usage. However, these two strategies are not always practical. Waiting for a fault to occur means the business is subject to regular downtimes and as discussed has huge potential negative consequences on the reputation of the product. Scheduled maintenance in other hand, can lower the down time and build up more reliability, however, it may not eliminate all the failures as unpredictable faults can still happen before the scheduled plan. Also doing maintenance on a set rate regardless of its actual condition means that most likely the equipment is still in a good condition at the time of the scheduled maintenance therefore the maintenance process could be unnecessary and wasteful.

The predictive maintenance is a modern approach as oppose to conventional maintenance plans. It suggests a forecast system that alerts the changes that leads to a failure in the system and gives a time window in which a preventive maintenance can and needs to be performed. The minimum benefits of a predictive maintenance system are as follow:

- Increasing the up time and reliability
- Minimizing the cost of maintenance by eliminating unnecessary repairs
- Preventing further damage caused by unforeseen failures

Nevertheless, predictive maintenance comes with its own difficulties mainly because it can be challenging to make accurate and reliable forecast for a complex system and it requires very careful approach to the data analysis [2, 5].

This paper aims to design and develop a Bayesian approach for predictive maintenance of complex and live systems by minimizing the costliest misclassification errors. The developed predictive model has to forecast the next failure of any complex and live system and to provide a time window to react to the upcoming faults for required preventive services. A turbofan jet engine is chosen as an industrial testbed in which the development and application of the predictive maintenance system has fully been demonstrated in the next sections.

II. METHODOLOGY

There are various classification approaches proposed and employed by numerous types of applications yet there is no one general algorithm to perfectly satisfy all kinds of application. In other words, the application and behavior of the data is vital in order to select a reliable predictor. Consequently, selecting an algorithm can be strenuous and tiresome and needs prior preparations.

Naïve Bayes' algorithm is one of the most versatile techniques in machine learning, which has attracted various academic research and industrial applications. Due to its simplicity, speed and reliability [5], using the Bayes theorem has been considered as a suitable choice for live monitoring, prognosis and diagnosis of complex system which require a predictive maintenance system. It proposes an equation to describe the probability of an event, based on prior evidence.

978-1-7281-3021-7/19/$31.00 ©2019 IEEE
DOI 10.1109/DeSE.2019.00181

Authorized licensed use limited to: UNIVERSITY OF CONNECTICUT. Downloaded on October 31,2022 at 21:37:03 UTC from IEEE Xplore. Restrictions apply.

Based on Bayes' statement, the likelihood of event Θ occurring, given that x is true is mathematically calculated as

$$P(\Theta|x) = \frac{P(x|\Theta)P(\Theta)}{P(x)} \tag{1}$$

where $P(\Theta|x)$ is the likelihood of event Θ occurring (given that x is true), $P(x|\Theta)$ is the likelihood of event x occurring (given that Θ is true), $P(\Theta)$ is the prior probability of event Θ and P(x) is the prior probability of event x. $P(\Theta)$and P(x) are the probabilities of event Θ and x occurring independently of each other and $P(x) \neq 0$ is called Marginal Probability. For classification purposes, Θ represents desired class group and x represent the data group. Next after description of the testbed, and the required data analysis and pre-processing, the implementation of Naïve Bayes Algorithm for classification is presented.

III. TURBOFAN JET ENGINE

Turbofan or Fan-jet is one of the most widely used jet engines in the Industry. Modern turbofans rely on several corresponding sensors which indicate the state of wear and the degradation of the engine. Turbo fan engines include 23 sensors that send live signals and report the most current state of each sub-section of the engine. This multidimensional system requires a robust algorithm in order to process each instance of the data and recognize the signs of the regression tear and wear in the device. Traditionally, every engine performs a certain number of flights (cycle) before a scheduled maintenance is performed regardless of the state of the engine. It is known that many of these engines are in a good condition when the procedure is performed so that those maintenances seem to be unnecessary [6].

A. Data and Data Preparation

The sensory datasets used in this paper are constructed by real-world datasets, which are provided by NASA [7] Prognostics Data Repository, includes 'Run to Failure' data for 100 different engines of the same type (Figure 1). In these datasets, each engine is assumed to start with different degrees of initial wear and manufacturing variations, which is unknown to the user. The datasets include the engine's IDs (Unit 1-100), the cycles since the observation (Time), 3 sets of settings (settings 1-3) beside various sensor readings. The time stamp for each engine starts from 1 and ends at the cycle when the failure occurs.

A number of sensory readings remain constant throughout or most of the observation time which mean they are not very useful for understanding the regression and wear of the engine. Since, most prediction algorithms struggle with complex systems and several predictors can have negative impacts on the accuracy, for this study all the constant values along with the 3 settings are removed. This has reduced the datasets to 14 main features.

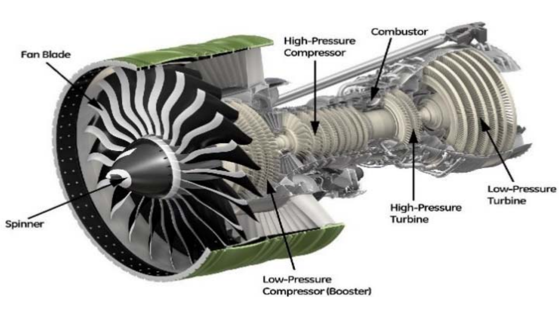

Figure 1. The type of turbofan that was used for 23 sensory readings [7]

Furthermore, the jet engine studies indicate that a level of noise is present within the data. It also showed a small number of outliers that located well away from the main data. This noise increases entropy and therefore can have a severe negative impact on the model efficiency [8]. Thus, a noise removal process was carried out on datasets using moving average technique. Outliers can also vastly impact on the 'mean' value and the 'standard deviation', the two principals of Bayesian theorem [8]. Consequently, outliers were identified and removed manually prior the application of Bayesian theorem.

B. Data Categorization

In order to predict a maintenance, the data was categorized to let the algorithm knows what the failure (output) looks like. Data cycles between 0 to 50 is labeled as 'Class1' or Red, which indicates an urgent state, i.e., an immediate maintenance is required. Data cycles between 51 to 100 are categorized as 'Class 2' or Green, which presents engines that require close observation continuously. Any other data is labelled as 'Class 3' or Blue meaning that the engine is in a healthy state.

In summary, the aim is to classify the health of the engine and indicates if the engine will require a maintenance within a defined period, i.e., 50 (Class 1 or Red), 100 (Class 2 or Green) or more (Class 3 or Blue) time stamps or flights.

IV. PREDIVTIVE MAINTENANCE

Naïve Bayes' theorem is one of the most fundamental probability strategies and despite the simplicity and speed, was able to compete with most crucial prediction methods within machine learning. The development of Bayes' Classifier using Gaussian normal distribution probability is presented here, where 14 identified attributes of datasets were used to categorise them according to 3 recognised classes for predictive maintenance.

A. Naïve Bayes Algorithm forNASA Jetengines

In a multidimensional system and based on Bayes' statement, if C is the class group and X is the data group:

C= {Red, Green, Blue}
X= $\{X_1, ..., X_{14}\}$

The probability that a given row data, X_i, belongs to class C_i is called the posterior of class C_i and is given by:

$$P(C_i|x_i) = \frac{P(x_i|C_i)P(C_i)}{P(x_i)} \tag{2}$$

In Naïve Bayes' algorithm, the decision as to which class is voted the most appropriate is based on the largest magnitude of posteriors associated to each class. In other words, the absolute value of calculated posteriors are not important but the relative sizes. All the class posteriors in the model have the same denominator (Marginal Likelihood) which means that $P(x_i)$ has no effect on the ratio and their relative magnitude and consequently has no effect on the Bayes' decision and can be safely disregarded. Therefore, the model equation becomes:

Authorized licensed use limited to: UNIVERSITY OF CONNECTICUT. Downloaded on October 31,2022 at 21:37:03 UTC from IEEE Xplore. Restrictions apply.

$$P(C_i|x_i) = P(x_i|C_i)P(C_i) \tag{3}$$

Predictor X is an array of 14 attributes. Therefore:

$$P(x_i|C_i) = \prod_{k=1}^{14} p(x_{i,k}|C_i) = p(x_{i,1}|C_i)p(x_{i,2}|C_i) \dots p(x_{i,14}|C_i)$$

The prior $P(C_i)$, is in fact the prior knowledge of the class C_i, based on previous experience. That is the probability that class C_i has occurred in the time of observation. Which is:

$$p(C_i) = \frac{n_{C_i}}{N} \tag{4}$$

where n_{C_i} is the number of time that class C_i was recognized and N is the number of training data rows. Therefore, the sum of the 3 priors of the model is equal to one, i.e.;

$$p(C_1) + p(C_2) + p(C_3) = 1 \tag{5}$$

The Naïve Bayes' classifier is constructed by determining the posterior of target class, i, and a comparison to posteriors associated to other classes in order to determine the class with the maximum value of posterior. Accordingly, each posterior term is calculated using the gaussian probability density:

$$P(x_i|C_i) = f(x_i|\,\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{6}$$

where μ is the mean of the distribution and σ is the standard deviation.

B. Implementation

The Naïve Bayes 'Classifier was implemented as a MATLAB program. The program requests the training and testing data. At first it splits the training data into three tables based on the class number. Then it calculates:

- The prior probability of each table (which is equal to prior of each class $P(C_i)$)
- The man value of each table (which is equal to μ_i for each class)
- The standard deviation value for each table (or σ_i in order to calculate the variance σ^2)

The Naïve Bayes' algorithm is based on voting for the highest posteriors in each instance of data. The posteriors are calculated by using gaussian density function (equation 6) for 10 datasets as an example for a randomly selected engine (Engine _ID=36) and are shown in Table 1, where all above three values used in the program and posteriors were placed in the table. Then each raw of the posterior table goes through a comparison process to determine the maximum posterior and thus the most likely class associated with each raw. The results were added to the fourth column of the table for future reference and will be compared with the test data's real labels and the accuracy was calculated as an output of the model.

The prediction made by the given Bayes' classifier resulted in 78.85% accuracy on a randomly selected engine (Number 36), which involved 156 raw of test data. Employing the posterior re-evaluation, prior to the use of cost matrix improved the overall accuracy from %78.87 to %90.03 [9]. It also minimized the serious fault from an initial value of %2.6. To break down the accuracy over each class, the confusion matrix of this prediction which involved 156 raw of test data was constructed and demonstrated in Figure 2.

TABLE 1. List of the calculated posteriors (scaled 0.01) for each class

	1	2	3
1	-0.265750425844130	-0.064091504491098	-0.057572837887655
2	-0.257823490506777	-0.058378185569937	-0.053263186482106
3	-0.262225275079254	-0.035461798962511	-0.022910786499715
4	-0.282201643993035	-0.060539711024468	-0.027340368549942
5	-0.256680460595495	-0.037739736994316	-0.034940695940658
6	-0.254069746653718	-0.035582810120059	-0.033261739462671
7	-0.261665062475777	-0.031069212171055	-0.016532021914451
8	-0.273659460131824	-0.045027657207139	-0.018148268340213
9	-0.240632666776873	-0.038698345895821	-0.054341788549852
10	-0.264963073288165	-0.043564001018080	-0.027851233591362

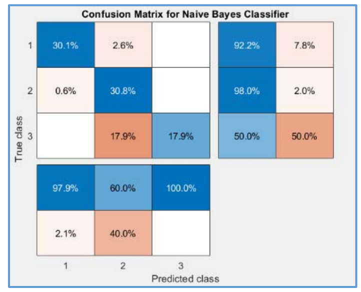

Figure 2. Results after smoothing and minimum cost technique

V. Conclusion

A predictive maintenance system was developed based on a Naïve Bayesian classifier and employed for prognostic and diagnosic of a turbofan jet engine as an industrial testbed. Bayesian theorem generally offers a robust method especially against noise and outliers while its simplicity and speed make it suitable for live monitoring algorithms. However, its sensitive to high entropy data meant that the structure of the data and application of the system were the main factors to deal with for having a quality model for the developed predictive maintenance system.

The employed industrial testbed was a complex system with noise contamination and almost uneven sensor readings. The success in obtaining good accuracy for this testbed indicates that the technique used for development of predictive maintenance system is adequate and can be used as a modern solution for reducing downtimes in complex systems such as airplanes' jet engines in comparison with other machine learning algorithms.

References

[1] The Royal Academy of Engineering, Industrial Systems: capturing value through manufacturing. (2012) https://www.raeng.org.uk /publications/reports/industrial_systems_capturing_value.

[2] A. Mosallam. Remaining useful life estimation of critical components based on Bayesian Approaches. Materials. Université de Franche-Comté, 2014. https://tel.archives-ouvertes.fr/tel-01412179

[3] The Manufacturer. (2018) https://www.themanufacturer.com/articles /machine-downtime-costs-uk-manufacturers-180bn-year/

[4] Maintenance procedures. (2018) http://www.hse.gov.uk/comah /sragtech/techmeasmaintena.htm

[5] J.V.Stone A Tutorial Introduction to Bayesian Analysis, Sebtel, 2013.

[6] S. Ackert. Engine Maintenance Concepts for Financiers, 2nd Edition, Aircraft Monitor, 2011.

[7] NASA, Prognostics Center - Data Repository. Available at: https: //ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

[8] T.J. Watson Research Centre (2018) An empirical study of the naive Bayes classifier. https://www.researchgate.net/publication /228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier.

[9] M. Saeidi. A Prognostic Bayesian Classifier, University of Salford 2018.

Authorized licensed use limited to: UNIVERSITY OF CONNECTICUT. Downloaded on October 31,2022 at 21:37:03 UTC from IEEE Xplore. Restrictions apply.