NRC Publications Archive Archives des publications du CNRC

Developing data mining-based prognostic models for CF-18 aircraft

Zaluski, Marvin; Létourneau, Sylvain; Bird, Jeffrey W (Jeff); Yang, Chunsheng

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1115/GT2010-22944>

ASME Turbo Expo 2010: Power for Land, Sea, and Air. Volume 3, Controls, Diagnostics and Instrumentation, pp. 235-243, 2010

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=d917439a-e8cb-4b1b-aa2b-f4075308fb2e>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=d917439a-e8cb-4b1b-aa2b-f4075308fb2e>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

GT2010-22944

DEVELOPING DATA MINING-BASED PROGNOSTIC MODELS FOR CF-18 AIRCRAFT

Marvin Zaluski

National Research Council of Canada
Ottawa, Ontario, Canada

Sylvain Létourneau

National Research Council of Canada
Ottawa, Ontario, Canada

Jeff Bird

National Research Council of Canada
Ottawa, Ontario, Canada

Chunsheng Yang

National Research Council of Canada
Ottawa, Ontario, Canada

ABSTRACT

The CF-18 aircraft is a complex system for which a variety of data are systematically being recorded: operational flight data from sensors and Built-In Test Equipment (BITE) and maintenance activities recorded by personnel. These data resources are stored and used within the operating organization but new analytical and statistical techniques and tools are being developed that could be applied to these data to benefit the organization. This paper investigates the utility of readily available CF-18 data to develop data mining-based models for prognostics and health management (PHM) systems. We introduce a generic data mining methodology developed to build prognostic models from operational and maintenance data and elaborate on challenges specific to the use of CF-18 data from the Canadian Forces. We focus on a number of key data mining tasks including: data gathering, information fusion, data pre-processing, model building, and evaluation. The solutions developed to address these tasks are described. A software tool developed to automate the model development process is also presented. Finally, the paper discusses preliminary results on the creation of models to predict F404 No. 4 Bearing and MFC (Main Fuel Control) failures on the CF-18.

INTRODUCTION

The CF-18 has been in service with the Canadian Forces (CF) since the early 1980's. The first aircraft was delivered in October 1982 and Canada's fleet logged more than 500,000 flying hours by 2002. At the time, the CF-18 was considered one of the most advanced platforms for diagnostics and the CF further improved the data management capabilities over the years. Aircraft operating parameters such as altitude, engine

exhaust temperature, shaft speeds, and airspeed are recorded throughout the flight and some parameters such as vibrations and pressures are recorded during a diagnostic event generated by the onboard Integrated Engine Condition Monitoring System (IECMS) BITE equipment. A diagnostic event signifies further investigation is required by the maintenance staff to rectify the problem. Diagnostic events are generated either when one or more parameter thresholds are exceeded, failure of one or more components, or manually triggered by the pilot. These operational data have been systematically archived by the CF and are available for diagnostic/prognostic purposes. In addition, a comprehensive database describing all maintenance actions performed by the personnel on the CF-18 is also available. GasTOPS, a control system and condition assessment company, and the Canadian Forces exploited some of these data to develop an advanced diagnostic/troubleshooting and performance monitoring tool [1]. In this paper, we investigate the potential usefulness of the available operational and maintenance CF18 data for prognostics.

The CF-18 maintenance database tracks over 2,300 components and each could be modeled using a data mining approach for diagnostics/prognostics. An initial investigation was performed using data from the No.4 Bearing and the resulting models showed potential for predicting component failure [2]. This paper describes a further investigation that focuses on the main fuel control (MFC) on the F404. The MFC is the hydro-mechanical component for numerous, complex engine control functions. It is a critical component since its failure can result in degraded performance, delays, cancellation of missions, and even possible loss of an engine or aircraft.

Prognostic models for the No.4 Bearing and MFC would enhance opportunistic and preventive maintenance procedures

with anticipated impacts on availability and cost. The objective is to build on previous experience in [3] and [5] with the application of data mining techniques for prognostics and to determine if the available CF-18 data could be used to build prognostic models for different components.

The next section offers a brief overview on our generic data mining-based approach to develop models for Prognostics and Health Management (PHM) systems and introduces a data mining tool that we are developing to automate and streamline the data mining process. We then elaborate on challenges specific to the development of prognostic models and detail the data mining experimentation performed. The paper presents preliminary results for both No.4 Bearing and MFC along with the evaluation criteria and some discussions.

GENERIC DATA MINING-BASED METHODOLOGY

Data mining-based prognosis is an emerging application of data-based modeling to real-world problems. The objective is to use data mining and machine learning techniques to build predictive models from readily available maintenance and operational databases. After validation, these models can be integrated into PHM systems to monitor equipment health and predict component failures before such events disrupt operations. Figure 1 illustrates the generic data mining methodology that we have developed to build prognostics models for complex systems such as trains [4] and aircraft [5]. This methodology consists of four steps: data gathering, data labeling, model evaluation, and model fusion.

Figure 1: Overview of Data Mining-Based Methodology

1. Data Gathering

Most data mining algorithms require a dataset as input, which contains instances consisting of vectors of attribute values. Modern machinery installations often generate many such datasets. Our first problem is to select the best dataset(s) to use to build models for a particular component. Advice from subject matter experts and reliable documentation can simplify this choice and help avoid a lengthy trial and error process. Not only must we select a dataset, we must also select a subset of instances to use in the analysis. The datasets are typically very large so it is inefficient to build models using all instances.

Simple solutions, such as random sampling, are also inappropriate. To build the desired predictive models, we must be much more focused. We base our analysis on data generated around each occurrence of failure (defined as component replacement). We first identify when replacements occurred and then retrieve instances around the time of these occurrences. Once we have the date and the part identifier for each replacement of a component, we retrieve the relevant instances from the selected sensor datasets. For each dataset and replacement, we retrieve the data obtained between m operating hours prior to the replacement and n operating hours after. The parameters m and n depend on the dataset and the component.

2. Data Labeling

In order to use supervised learning algorithms, we need to add a class attribute (or label) to the selected sensor data. We propose an automated labeling approach. This approach labels as positive all instances that fall in a pre-determined target window before the occurrence of a component failure and as negative all other instances. This labeling scheme allows us to build a classifier that generates an alert whenever the patterns in the data are similar to the ones observed near a failure. In practice, we define the length of the target window by taking into account the optimal period for the end users to receive the alerts and the balance between positive and negative instances. In practice, the proportion of positive instance should be over 15%.

3. Modeling and Evaluation of Models

The modeling and evaluation stage starts with a systematic investigation of data transformations that can be used to improve the initial, as measured, representation. This is done by augmenting the initial representation with new parameters or features created using methods from process physics, signal processing, time-series analysis, and constructive induction. Feature selection is also applied on the augmented data representation to automatically remove redundant or irrelevant features [6][7]. After updating the initial dataset with the class attribute and incorporating data representation enhancements, we use supervised learning algorithms to generate the prognostic models. We use data from a subset of all failures for developing the models and keep the remaining data for testing. Any supervised learning algorithm can be used. In early experiments, we tend to prefer simple algorithms such as decision trees and Naive-Bayes over more complex ones because of their efficiency and because they produce models that we can easily explain to the end users. We apply the same algorithm several times with varying attribute subsets and optimization of objective information. To compare the models obtained, we apply a score-based approach [2, 3] that we have developed to evaluate models for prognostic systems.

4. Model Fusion

Model fusion can be used for two reasons. First, when more than one dataset is relevant for a given component, we can build a model for each dataset and then use model fusion to combine predictions from the various models. Second, we can apply model fusion for performance optimization regardless of the number of datasets selected. In this case, we learn various models using various techniques or parameter settings and combine them to obtain better performance than using any single model. Bagging and boosting [8] are two popular techniques to combine models but they are only applicable when there is a single dataset and one kind of model (a single learning algorithm). For heterogeneous models or multiple datasets, we apply methods based on a voting or stacking strategy [9]. These techniques are globally referred to as multiple classifier systems [10].

AUTOMATION OF DATA MINING PROCESSES

In order to obtain accurate prognostic models, we often need to iterate over many of the steps in the data mining-based methodology presented above. For instance, in most experiments, we need to repeat the execution of the same software module with different configurations or input parameters. To facilitate the process, we have developed an automated data mining system named EBM3 (Environment for Building Models for Machinery Maintenance). In addition to speeding up the execution, the EBM3 system allows the research team to maximize reuse of software components and experimental methodologies between applications.

EBM3 provides a simple user interface that the researchers use to define the overall data mining experiment that needs to be executed to build the desired prognostic models. As mentioned above, this experiment involved a number of tasks such as data pre-processing, modeling, evaluation, and model fusion. For each of these tasks, the researcher specifies the input, output, task specific parameters, and the actual software component that will perform the required computation. These software components could be written in various languages including Java, Perl, SAS, R, C/C++, Python, Jython, Jess, and Matlab. They can also be written for a specific operating system (Windows or Linux) or specified to run on a given computer. During the execution, the EBM3 system automatically launches the various software components. As needed, it automatically connects to the specified computers (using the provided computer IP and login information). The data are automatically transferred from one step to the following one with transparent data format conversions as required. Specialized conversion programs were developed and integrated into EBM3. The output format from a process is compared to the input format for the next process. If the data formats are not the same, an applicable conversion program is identified and executed to transform the output format into the input format. Mechanisms to monitor the execution and to handle run-time errors are also implemented within EBM3.

EBM3 supports the concept of iterations to allow a systematic exploration of parameter values as well as

repetitions for various datasets and learning algorithms. Iterations can be defined over a list of discrete values (e.g., algorithm and dataset names) or with a numeric range with predetermined increments. This functionality is particularly useful when exploring the potential of various learning algorithms and data transformations. The EBM3 system stores the results for all iterations, which allows detailed analysis of the results.

Another important feature of EBM3 is its ability to automatically deploy to users the completed predictive models. After the desired model(s) are developed, EBM3 encapsulates all the required software and configuration information into a package that is ready to be deployed within the Java 2 Platform Enterprise Edition (J2EE) infrastructure that we are using to implement delivery of real-time PHM functionalities. This provides seamless integration between building and deploying models in developing PHM systems.

In sum, the EBM3 system provides a structured framework for the definition and execution of complex data mining processes where each step can be implemented independently with great flexibility in the choice of programming languages, tools, and computer environments. Advanced features allow efficient execution of repetitive tasks and deployment of the models built. As discussed throughout the paper, the EBM3 system has been used extensively to investigate the potential use of CF-18 sensor and maintenance data for prognostics.

CHALLENGES

Some prognostics challenges are the same for both No.4 Bearing and MFC and these challenges are outlined in [2]. The challenges in implementing the generic data mining methodology for the CF-18 can be separated into two different groups. The first group of challenges is related to transforming the initial sensor measurements into a format that can be used for data mining. Specific difficulties include data selection, parameter creation, feature selection, handling of missing data, different recording frequencies, and conversion of the operational data to useable formats. Another group of challenges is related to implementing the generic data mining methodology. These challenges include algorithm selection, feature dataset selection, evaluation criteria, and adaptation of these tools and processes into an automated EBM3 data mining environment. The optimal solution would implement a data mining model with very high precision and accuracy while minimizing errors. Since the current objective was not to find an optimal solution, we only considered a limited number of approaches in this initial investigation.

The CF-18 data made available for this research consist of five years of operational sensor data (from 1999 to 2004) and over ten years of maintenance data. The sensor data were stored in over 25,000 binary flat files call Aircraft Data Files (ADF) while the maintenance data were in an Oracle database format. An analysis of the maintenance database revealed 1280 cases of No.4 Bearing replacements and over 1,700 cases of MFC replacements from 1999-2004. The maintenance database

established that a bearing is no longer useful when its inventory status is set to scrapped or no longer serviceable. The same strategy of determining if an MFC is no longer useful does not apply because this component can be rebuilt and recertified for service. This rebuilt and recertified component can be identified and tracked along with other major F404 components such as the high pressure turbine and fan. Therefore, the retrieval of relevant maintenance and operational data is different depending on the type of component.

The paper only considers component replacements that occurred during the time for which operational data were available (1999-2004). Also, to remove some variability related to aircraft configuration, we only retained replacements that stayed on the same engine, same aircraft, and same position for all their life. We obtained 42 examples of No.4 bearing replacements and 6 examples of MFC replacements that satisfy these restrictions. Future work might look at a less restrictive set.

The engine, aircraft, and engine position (left or right) information was not explicitly recorded in the maintenance database, but this information could be traced by joining related engine and aircraft maintenance actions to the bearing replacement. A special application was developed and integrated in EBM3 to perform this analysis as part of the data gathering task in the data mining experimentation described below. An example of this integration of maintenance information is described for MFC BECA0527 using date information and next higher component information stored with each maintenance record. MFC BECA0527 was initially installed in April 2001 on Engine 376045. This engine was installed on aircraft 719 on the right hand side on June 2000. Since the MFC can be replaced without engine removal, the operational data recorded after the date of installation of the MFC, April 2001, will be used in the data mining experiment. A similar verification is done for the removal of MFC BECA0527 to ensure the same configuration, engine, aircraft, and position was maintained.

The operational data collected on the CF-18 aircraft are organized into 65 different message groupings. Each of these message groupings focus on a specific function of the aircraft such as armament, engine performance, and diagnostic information. We focused on four of these to retrieve relevant sensor measurements and flight information because they were most relevant to predicting component problems in the engine. One problem was establishing when a flight started and ended so that the information could be gathered. Another problem was the differences in recording frequencies between parameters; some parameters were recorded every second while others every five seconds during flight. Examples of such parameters are air speed, low pressure rotor speeds, and exhaust gas temperatures. Other parameters, such as oil pressures and broadband vibration, are recorded once per second for 15 seconds during a periodic thrust check. These parameters are also recorded once per second for 35 seconds after a diagnostic event is triggered by the Built In Test Equipment (BITE). These

different sampling rates result in different amounts of data for each sensor during the flight and many model building algorithms do not handle varying data set sizes for a single flight. It was important to summarize these differing operational data records into a single value to be used in model development because all of these parametric values will be considered in building prognostic models for the component. For instance, the parametric data for the flight can be summarized into statistical values such as an average, median, minimum, or maximum. Since there are many components and symptoms that could be modeled using our data mining techniques, we wanted to focus on components tracked in the maintenance data where the components could be replaced without engine removal to remove this source of performance variability. For these and other reasons, we selected the MFC as a feasible and important component for PHM. By building prognostic models for this component, we hope to predict failure/replacement of the MFC before it actually fails and causes an unexpected disruption and/or cost to the operator. If this impending failure/replacement can be predicted, this would allow the CF to perform more opportunistic maintenance, reduce secondary damage and improve overall CF-18 fleet readiness.

PROGNOSTIC MODEL DEVELOPMENT

This section presents the experiments for developing prognostic models for No. 4 Bearing and MFC replacements from the archived CF-18 operational and maintenance data. The experiments were conducted following the generic methodology described, with the EBM3 system. This section describes the details for each of the different stages in the methodology and how the indicated challenges are addressed.

The data mining experimentation used the EBM3 functionality to implement the generic data mining methodology. EBM3 uses eXtensible Markup Language (XML) to design and store data mining experiments, and integrates external applications. These external applications are called blackboxes, which are developed to execute an independent data mining task in a specific environment or platform. The CF-18 component experimentation used over 30 external applications or blackboxes from EBM3 to perform the different tasks required in each stage of the generic data mining methodology. There were also over 85 XML experiment files created for each CF-18 component to separate and balance the computing load for building datasets and models. The experiment benefited from EBM3's ability to automate the iterations over different feature sets and machine learning algorithms. The other XML experimentation files processed the results from the evaluation stage by generating summary HTML files. The CF-18 No.4 bearing and MFC experimentation used three of the four stages in the generic data mining methodology and they are further explained below. In this work, we have not conducted any experimentation related to the last stage: model fusion.

1. Data Gathering and Labeling

The data gathering and labeling processes transform the initial sensor measurements into a format that a machine learning algorithm can use to build a data mining model. Figure 2 illustrates the processes involved in generating the data set files from the raw sensor data. The original format of the operational data was binary flat file that was processed by a Java application into an Oracle database, which allows for easy retrieval. The next step is identification of the appropriate operational data using the component replacement data from the maintenance database. For the No.4 Bearing, the maintenance data were identified when the bearing was initially installed, and when it was removed with no further useful life remaining. For the MFC, a different method was used to identify MFC problems in the maintenance data that it identified a problem whenever there was an installation and removal of a MFC.

A Java application retrieves the information for installation and removal from the maintenance database. This maintenance information is used to collect the operational data and the same Java application transforms it into a text flat file with raw operational data. This flat file also has each observation labeled as Normal Operation and Alerting Period added to it. The labeling of the data for the MFC was different from the No.4 Bearing [2]. The Alerting Period was determined as being the last 20 flights before removal of the MFC. The Normal Operation Period includes all flights that occur after the installation of the MFC up to the last 20 flights before the MFC was removed. Each MFC replacement generated a flight data file. Accordingly, 6 Flight Data files were generated during this step to cover the 6 MFC removals. The amount of data in each of these data files ranged from 32 flights (61 hours) to 343 flights (640 hours).

The next step for both No.4 Bearing and MFC data gathering involves 21 different Java applications to summarize the numerous data points for a single flight parameter into one statistical value for the flight parameter. Traditional statistics such as average and standard deviation are calculated. Also, some ordered statistical values are calculated such as medians and quantiles. This process results in 21 different datasets with summarized CF-18 operational flight data for the different statistical measures. Some of the features in the CF-18 operational flight data sets must be filtered to remove noise or bias. An EBM3 Experiment was created to filter out such attributes as flight number, date information, and aircraft configuration information. Finally the same EBM3 Experiment separates the CF-18 operational flight data sets into training and testing data sets. A Leave One Batch Out (LOBO) cross validation approach was implemented for the MFC experimentation so the splitting of the data into training and testing differed from that of the No.4 Bearing in [2].

The MFC replacements were ordered by Problem ID (in collected relevant dataset, there are many time series which are associated with failures. To distinguish each time series, we add a unique problem ID to each series.) and the one Problem ID

was selected for testing data. All of the data related to the Problem ID for testing were placed in a testing data file. The other Problem ID operational data were used for training and used to create a training data file. Each statistical summary feature resulted in 6 different training and testing data sets reflecting the 6 different identified MFC problems. The end of the data gathering and labeling phase resulted in 126 training and 126 testing data sets, one for each of the 21 statistical summary feature and Problem ID. All of these 126 training and 126 testing data sets are used in the model building and testing process.

Figure 2: Data Gathering and Labeling Processes in EBM3

Algorithm Type	Algorithm Implementations	No.4 Bearing Model	MFC Model
Bayes Learners	AODE	X	X
	ComplementNaiveBayes	√	X
	NaiveBayes	√	√
Lazy Learners	IBk	√	√
	KStar	√	√
Rule Learners	ConjunctiveRule	√	√
	JRip	√	√
	NNge	√	√
	OneR	√	√
	Ridor	√	√
	ZeroR	√	√
Tree Learners	J48	√	√
	ADTree	√	√
	DecisionStump	√	√
	Id3	X	X
	M5P	X	X
	NBTree	√	√

Table 1: Machine Learning Algorithms Used

2. Model Building

EBM3 integrates the Weka suite of machine learning algorithms. The Weka suite of machine learning algorithms

contains many different machine learning algorithms for classification [11] and EBM3 integrates each of these algorithms into its data mining environment. In the CF-18 component data mining experiment, four different types of classification algorithms were used. These algorithms used are outlined in Table 1. Not all algorithms were successful in creating data mining models and those which were successful in generating a model are also identified in Table 1.

EBM3 experiments were developed to generate models from the 126 different training data sets from the four types of machine learning algorithms. Figure 3 below outlines the EBM3 experiments and the models that were generated. A total of 2,288 models were created after executing the EBM3 experimentation.

Figure 3: EBM3 Experimentation in Model Building

3. Model Evaluation

There were many different processes involved in calculating and displaying the results of model evaluation. An EBM3 experiment was created to integrate these different processes and further automate the model evaluation phase. The first process calculates the different evaluation measures and creates a flat text file. This flat text file is used to create HTML pages and supporting graphics that display the results. These processes are described in Figure 4 below.

Figure 4: EBM3 Experimentation Model Evaluation

EXPERIMENTAL RESULTS

1. Evaluation Criteria

Two different types of evaluation criteria are used to evaluate the models built. The first set of evaluation criteria used are traditional techniques for evaluating data mining models. Precision and recall are two measures used to assess the performance of a model [12]. Precision is calculated using the formula below and is the total number of correct predictions over the total number of predictions made. Recall is the total number of correctly predicted replacement alerts over the total number of replacement alerts. A replacement alert identifies that the data model predicts a component failure and the component should be replaced at the discretion of the maintenance technician.

$$\text{Precision} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Formula 1: Precision Formula (I think the format is just [1] and the end of the line?)

$$\text{Recall} = \frac{\text{Number of Correct Replacement Alert Predictions}}{\text{Total Number of Replacement Alerts}}$$

Formula 2: Recall Formula

In addition, we also relied on two custom metrics to take into account the timeliness of the alerts produced. The first metric, named the problem detection rate, indicates the proportion of problems that have been detected during the target detection period. The second metric, named score, is tailored from related work on the application of the proposed data mining methodology to predict failures in the railway and commercial aerospace domains [3, 4]. The score metric uses a custom reward function that assigns a value to each prediction from the model based on the time difference between generation of the failure prediction and the actual failure. As shown in Figure 7, maximal rewards are obtained in the target detection period. The final score for a given model is obtained by adding the rewards associated with the alerts produced by this model while running it on the test dataset.

Figure 7 Scoring Function

These four evaluation measures are applied to the predictions from the models generated for both CF-18 No.4 Bearing and MFC replacement.

2. Evaluation and results

Since the MFC experimentation uses cross validation, we needed to combine results from the various test datasets (various iterations of the cross-validation procedure) to obtain the overall measure of performance¹. Since each Problem ID was used once for testing, there were 6 different prediction files generated for each statistical summary feature. In total there were 2,288 experiment runs. For each run, we obtained a result file. Therefore, there were 2,288 different result files, one for each of the different algorithms, problem IDs, and feature sets. These flat files were combined and used to calculate the four different evaluation metrics for each of the statistical summary features. The total number of flights in the operational data was 798 and the number of Alert Period labeled flights was 120, Normal Operation was 678. The flight operational data set has an unequal number of normal and alerting situation examples and this will bias the generation of the prognostic models. Table 2 outlines the results of these calculations for the best performing algorithms and feature datasets for No.4 Bearing experimentation. Three different algorithm results are presented in Table 2, they are: Conjunctive Rule, Nearest Neighbor generalized exemplar (NNge), and Naïve Bayes Decision Tree (NBTree). The Feature Data Sets outlined in Table 1 and 2 are ordered statistical values and these values are calculated by ordering the raw data into a list of ascending values then identifying specific quantile values and positions such as 10th, 25th, 75th, and 90th quantiles. The Interquantile Average 10 90 feature dataset is the average of values between the identified 10th and 90th quantiles. Similarly the Upper Quantile Average 75 is the average of the values greater than the 75th quantile.

Feature Data Set	Algorithm	Performance Measures			
		Score	Problem Detection Rate	Precision	Recall
Interquantile Average 10 90	Conjunctive Rule	-4	33%	51%	8%
	NNge	33	100%	61%	72%
	NBTree	49	100%	61%	68%
Upper Quantile Average 75	Conjunctive Rule	-200	100%	49%	87%
	NNge	73	100%	59%	44%
	NBTree	-18	100%	56%	58%

Table 2: Preliminary Results for No.4 Bearing Replacement

¹ Please note that cross-validation was not used in our previously reported results for the No.4 Bearing [2].

The data mining models detected all the No.4 Bearing replacements except for the model created with the Conjunctive Rule algorithm on the Interquantile Average 10 90 data set. The best score obtained is 73 for NNge on the Upper Quantile Average 75 dataset where the best potential score is 675. NBTree and NNge have the highest precision, 61%, on the Interquantile Average 10 90 data set. Finally, the ConjunctiveRule algorithm on the Upper Quantile Average 75 dataset outperformed the others with a recall of 87%.

Feature Data Set	Algorithms	Performance Measures			
		Score	Problem Detection Rate	Precision	Recall
Inter-quantile Average 10 90	NNge	-1276	67%	54%	20%
	NBTree	-1272	67%	54%	22%
Lower Quantile Average 10	NaiveBayes	-910	83%	63%	23%
	SMO	15	33%	85%	5%
Inter-quantile Average 25 75	NaiveBayes	-2243	100%	31%	62%
	SMO	-1178	83%	57%	28%

Table 3: Preliminary Results for MFC Replacement

Similarly, Table 3 outlines the best performing algorithms and flight feature summary datasets for the MFC component. We first noticed that the NNge models which obtained very good performance on the No.4 Bearing did not perform as well for the MFC component. Instead, our investigation suggests that the most promising combinations of feature sets and algorithms for the MFC are as follows. The NaiveBayes algorithm using the Interquantile Average 25 75 feature set was able to detect all problems in the MFC data and yield the best recall metric of 62%. The Lower Quantile Average 10 and SMO algorithm had the best score, 15, and highest precision, 85%. The Lower Quantile Average 10 is the average of the values below the 10th quantile and the Interquantile Average 25 75 is the average of values between the 25th and 75th quantiles.

DISCUSSION

To summarize, desirable models could be selected from those that were developed from operational and maintenance data that offered the following encouraging performance:

- No. 4 Bearing: 61% precision-failure forecast and 87% of the alerts for predicting need for replacement
- MFC: 85% precision-failure forecast and 62% of the alerts for predicting need for replacement.

The experimentation results for the MFC and No.4 Bearing varied significantly among the models and datasets. Good

results achieved in modeling the No. 4 Bearing replacement using a certain modeling technique and a feature data set did not guarantee the same model performance when using the same modeling configuration for MFC replacement. This can be attributed to many different issues which we briefly discuss below. These issues confirm that a multi-model approach must be taken to determine the viability of a model fusion.

First, the quantity of the data available for the No.4 Bearing and the MFC components was quite different. The maintenance and operational data related to No.4 Bearing had more examples of replacements, 42, than the MFC, 6. The small number of MFC replacement cases forced us to use a cross-validation approach to evaluate the expected performance while this was not required for the No.4 Bearing. A higher number of MFC replacement examples would be highly desirable. One possible solution would be to relax the constraints on replacement identification in the maintenance data. For example, we could include MFC components that may have been removed, but their installation happened before the operational data we received. Another issue with the MFC datasets was that they did not have an equal number of normal operation and alert situation observations which created an imbalanced learning problem. This has an effect on the quality of models built from the data mining process because the algorithms are sensitive to this imbalance. Some techniques that could be used are over sampling and under sampling. Over sampling is defined as adding additional observations from the smaller class population. Under sampling is defined as removing observations from the larger class population. Each of these techniques helps in balancing the dataset, but they could also introduce new issues like which observations to leave out or include.

Second, the dataset quality differs for the No.4 Bearing and MFC. The No.4 Bearing dataset contains only the first 20 flights after initial installation and the 20 flights before final removal. The MFC data contained all the flight data from installation to removal. This made the task of identifying problems more challenging especially as the operational data reached the alerting period, but was still classified as normal operation. Also, this resulted in different numbers of flights and amount of operational data for each MFC replacement. Finally, any erroneous flight data due to recording was filtered out for the MFC data set, while the No. 4 Bearing contained these erroneous data.

Third, some of the evaluation measures used were traditional machine learning metrics such as precision and recall. These may not be the best measures for assessing data mining-based model performance. Other custom evaluation measures were introduced to address this issue, but other evaluation metrics may need to be developed such as alert precision, latency and assessment of false alerts. These new evaluation metrics may require additional data and these data would need to be retrieved from the maintenance and operational data sources.

Also, the operational data generated by the aircraft have over 180 different parameters and some of the parameters are likely not directly related to component failure or engine performance. For instance, aircraft navigational information may not be needed for building data mining-based prognostic models and these parameters may hinder performance of the prognostic models through the introduction of noise in the data. Since some data mining algorithms are sensitive to noise in the data, it can be beneficial to filter out this extra information from the datasets used in learning prognostic models. One method of feature filtering is the use of domain expert knowledge to determine which operational parameters are influential in characterizing an impending failure. Another method could be the use of feature evaluation applications within Weka to empirically evaluate each of the operational data and provide a ranking of the most influential parameters. This list of ranked parameters could be used to construct filtered operational data sets to be used in the training of data mining-based models.

Feature engineering can also be used to improve the performance of the data mining-based models through the introduction of new physics model-based features that would be derived from the aircraft's usual recorded parameters. For instance, it would be interesting to use temperature corrected rotor speeds, the ratio of engine rotor speeds or the fuel ratio units (fuel flow/compressor pressure) as these correspond to MFC functions and control parameters. Other physics -based, hybrid models may improve data mining model performance. The consequence of using engineering features is that some limited, additional preprocessing of the data will need to be done in a real-time deployed prognostic system.

Finally, the data mining-based model building approach must handle the different aircraft component characteristics. For instance the No.4 Bearing when damaged is not able to be rebuilt as is the case with the MFC. Also, each aircraft component has different failure modes and each failure mode has different probabilities of occurrence given the number of engine operating hours. This information can be obtained from the FMEA (Failure Mode and Effects Analysis) documentation. This information could be used to select parameters and to group the different component replacements into similar, common cause or cascading failure modes. Such grouping could be used to build separate data mining-based models for predicting specific component failure modes. Alternatively, the FMEA documentation could be used to build other knowledge bases in other reasoning systems such as Case-Based Reasoning in [13] which could later be integrated with data mining to help improve the performance of the prognostic models.

CONCLUSION AND FUTURE WORK

This paper describes a generic data mining methodology and how it is applied to predicting replacement of the No.4 Bearing and MFC on the CF-18. The already available

maintenance and operational data resources are used with the methodology to build and evaluate predictive data mining models, which can be used in a PHM system. The preliminary results show that the developed prognostic models are able to predict a significant proportion of No.4 Bearing failures using data resources readily available within the Canadian Forces. Other potential areas of future work could be feature selection, feature engineering, hybrid models, model fusion and more accurate predictions of remaining useful life. These and other types of experimentation are ongoing and will be combined with improved evaluation criteria to ensure that scoring metrics match the operator's needs. Finally, we intend to implement a demonstration PHM system with selected data mining models to help further assess the potential benefits of PHM for the CF-18.

ACKNOWLEDGMENTS

This work was made possible with the collaboration and contributions from Canadian Department of National Defense and Defense Research and Development Canada.

Many people at the National Research Council have contributed to this project. Special thanks go to Bob Orchard, Chris Drummond, Elizabeth Scarlett, Richard Liu, George Forester, and Kyle Mulligan.

REFERENCES

1. Cue, R. W. and Muir, D. E., "Engine Performance Monitoring and Troubleshooting Techniques for the CF-18 Aircraft", *Journal of Engineering for Gas Turbines and Power*, vol. 113, pp. 11-19, January 1991.
2. M. Zaluski, S. Létourneau, C. Yang, "Data Mining-Based Prognostics for CF-18 Components," in *Proceedings of the 13th CASI Aeronautics Conference (AERO 2009)*, Kanata, Ontario, Canada, May 5 – 7, 2009.
3. S. Létourneau, C. Yang, C. Drummond, E. Scarlett, J. Valdes, and M. Zaluski, "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.
4. C. Yang and S. Létourneau, "Learning to predict train wheel failures", in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2005)*, Chicago, IL, USA, August 2005, pp. 516–525.
5. S. Létourneau, A. F. Famili, and S. Matwin, "Data mining for prediction of aircraft component replacement," *IEEE Intelligent Systems and their Applications*, vol. 14, no. 6, pp. 59–66, 1999.
6. M. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of the 17th International Conference on Machine Learning*, Aug 2000, pp. 359–366.
7. K. Kira and L. Rendell, "A practical approach to feature selection," in *Proceedings of the 9th International Conference on Machine Learning*, 1992, pp. 249–256.
8. T. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization," *IEEE Intelligent Systems and their Applications*, vol. 40, pp. 139–158, 2000.
9. I. K. G. Tsoumakas and I. Blahavas, "Effective voting of heterogeneous classifiers," in *Proceedings of the 15th European Conference on Machine Learning (MCML2004)*, 2004, pp. 465–476.
10. S. Dzeroski and B. Zenko, "Stacking with multi-response model trees," in *The Proceeding of International Workshop in Multiple Classifier Systems (MCS2002)*, 2002, pp. 201–211.
11. Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques," 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
12. Makhoul, J., Kubala, F., Schwartz, R., and Weischedel, R. (1999), "Performance Measures for Information Extraction," in *Proceedings of the DARPA Broadcast News Workshop*, Herndon, USA, February 28 – March 3, 1999.
13. Atamer, A. (2004), "Comparison of FMEA and field-experience for a turbofan engine with application to case based reasoning," in *Proceedings of the IEEE Aerospace Conference*, 2004.