A CONSTRAINT-BASED APPROACH TO PREDICTIVE MAINTENANCE MODEL DEVELOPMENT

Joe Gorman, Glenn Takata, Subhash Patel, and Dan Grecu
Charles River Analytics

ABSTRACT

Predictive maintenance is the combination of inspection and data analysis to perform maintenance when the need is indicated by unit performance. Significant cost savings are possible while preserving a high level of system performance and readiness. Identifying predictors of maintenance conditions requires expert knowledge and the ability to process large data sets. This paper describes a novel use of constraint-based data-mining to model exceedence conditions. The approach extends the extract, transformation, and load process with domain aggregate approximation to encode expert knowledge. A data-mining workbench enables an expert to pose hypotheses that constrain a multivariate data-mining process.

KEY WORDS

Predictive Maintenance, data mining, constraint-based data mining, domain aggregate approximation, flight safety

INTRODUCTION

Predictive maintenance is the combination of inspection and data analysis to perform maintenance when the need is indicated by unit performance. The method can provide significant cost savings to operators while preserving a high level of system performance and readiness. The monitoring of complex instrumented systems creates vast collections of observations. Identifying predictors of maintenance requires expert knowledge and the ability to process large data sets. This paper describes a novel application of constraint-based data mining to model exceedence conditions in twin engine aircraft. The approach extends the extract, transformation, and load (ETL) process with domain aggregate approximation (DAX) to encode expert knowledge in the transformation step. A data mining workbench enables a human expert to pose hypotheses that constrain a multivariate data mining process.

Air and space systems have long been extensively instrumented. The availability of performance measures enables a range of safety and cost benefits (Wilmering & Ramesh, 2005). However, the volume of data and the complexity of relationships (Williams, 2006) have limited the industries ability to achieve improvements in flight safety and cost reductions. This paper describes a system for building condition models from the combination of expert knowledge and constraint-based data mining on digital flight data records (DFDR). The next sections discuss condition-based maintenance and constraint-based analysis. The following sections describe a research prototype for Constraint-based Analysis for Flight Operations (CBA-OPS) that implementes constraint-based analysis to characterize safety relevant conditions. The paper concludes with a discussion of CBA-OPS performance and recommends extensions and future work.

CONDITION-BASED MAINTENANCE

The safe operation of twin engine aircraft depends on routine maintenance and expensive periodic engine rebuilds. Work on the Space Shuttle Main Engine (Malloy et al., 1997) established a foundation for the computer analysis of rocket motors. Letourneau et al. (1999) proposed the use of data mining to predict component failure in aircraft.

Data mining is an effective machine learning technique for extraction information from large datasets (Han and Kamber, 2000). However, the size of DFDR data sets remains beyond the effective processing capacity of data mining algorithms (Keller et al., 2001). The multi-variant character of DFDR data further complicates data mining (Oats, 1999).

CONSTRAINT-BASED ANALYSIS

Data mining is typically performed on data warehouses that are optimized for analysis (Han and Kamber, 2000). Vlachos et al. (2003) critique several data compression strategies for time series data while Lin et al. (2003) presents a procedure for symbolic aggregate approximation that both compresses the data set and prepares it for frequent pattern mining. CBA-OPS extends this process by encoding expert domain knowledge through the aggregate approximation. The domain aggregate approximation (DAX) of DFDR data is based on time domain features described by Zakrajesik, Fulton, and Meyer (1994) and is parameterized to compensate for the varied character of each data series.

DAX compresses the data set but does not change the dimensionality of the search space. The well known apriori principle allows a data-mining engine to trim the search space dynamically but must consider all alternatives in the early stages of the search (Agrawal & Srikant, 1995). Constraint-based data mining combines initial user knowledge and the apriori principle to reduce the dimensionality of the initial search space (Han, Lakshmanan, & Ng, 1999).

CBA-OPS

CBA-OPS demonstrate constraint-based analysis of aircraft operations in general and engine and vehicle health maintenance operations in particular. A parameterized implementation of DAX is performed during data loading. A graphical constraint definition language (CDL) is used to encode expert knowledge in the form of hypotheses that describe elements of the pattern that is being sought. The apriori principle permits reduction of the search space to only those patterns that contain the hypotheses (Han, Lakshmanan, & Ng, 1999). Automated frequent pattern mining proceeds from this strong foundation using the apriori algorithm to dynamically limit the search space as well. The combination of expert knowledge and machine learning permits the discovery of complex patterns in multivariate data.

SYSTEM ARCHITECTURE

A full function proof of concept prototype of CBA-OPS was implemented. The functional architecture of the prototype is shown in Figure 1. The Data Generation Component provides an interface to data sources. Interfaces to both a flight simulator and a representative DFDR source were implemented. DAX processing was implemented as part of the data validation and data cleaning operations in the ETL component.

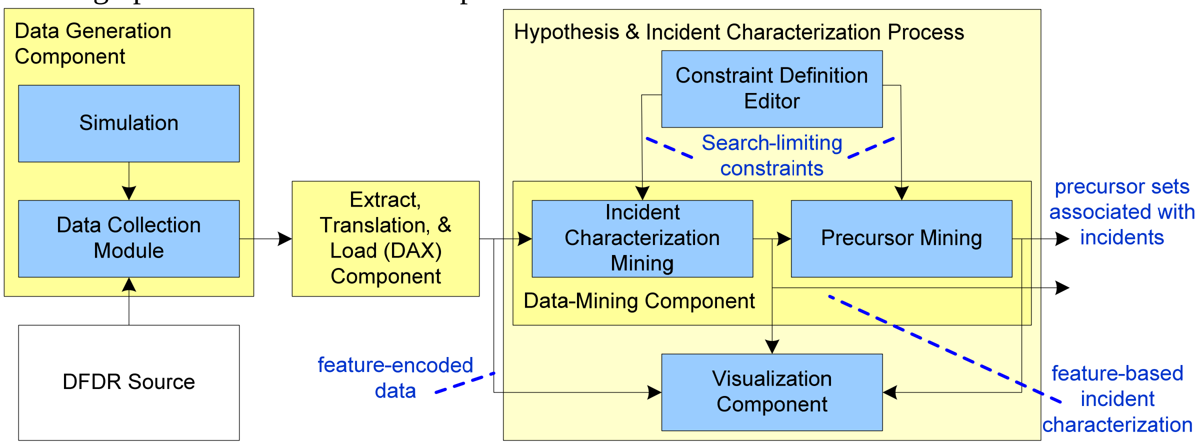


Figure 1: CBA-OPS Functional Architecture

The CBA-OPS prototype is incident oriented. The original research objective was to characterize flight safety relevant conditions. A representative scenario was developed base upon the analysis of Intersun Sunways Flight 2042 on 1 October 1996 (Flight Safety Foundation, 1998). The Constraint Definition Editor (CDE) provides a convenient graphical interface for domain experts to form hypothesis about a safety relevant incident. The CDE interfaces with the Data-Mining Component to expand the initial hypotheses into a characterization of the incident. The Visualization Component provides views of feature-encoded DFDR source data that are selected either by using a data browser or as the results of data-mining operations.

ETL AND DAX

The ETL process data validation and cleaning before feature encoding. Encoding is a two step process. The first is a piecewise segmentation of each signal. The duration of segments is variable and determined by the error in the linear approximation. A custom de-noising algorithm was implemented to manage the impact of noise in the source on the segmented form. A fusion algorithm attempts to fuse a noisy segment with preceding or following segments before the segment is added into the segment stream.

After the end points of a segment are determined the segment is subdivided to maintain a close correlation to the original data as is illustrated in Figure 2. Both the error in the linear fit and the number of segment subdivisions are user specified parameters. Segmentation alone was found to provide a 4 to 12 times reduction in the size of the data set.

Domain aggregate approximation (DAX) is performed in the second step. DAX extends the approach described by Lin et al. (2003) by explicitly expressing expert knowledge in the symbol definitions. Following Zakrajsek, Fulton and Meyer (1994) defines a vocabulary of eight parameterized features that are observed in the time domain trace of a signal. CBA-OPS experienced up to a 90:1 data compression ratio after segmentation and DAX.

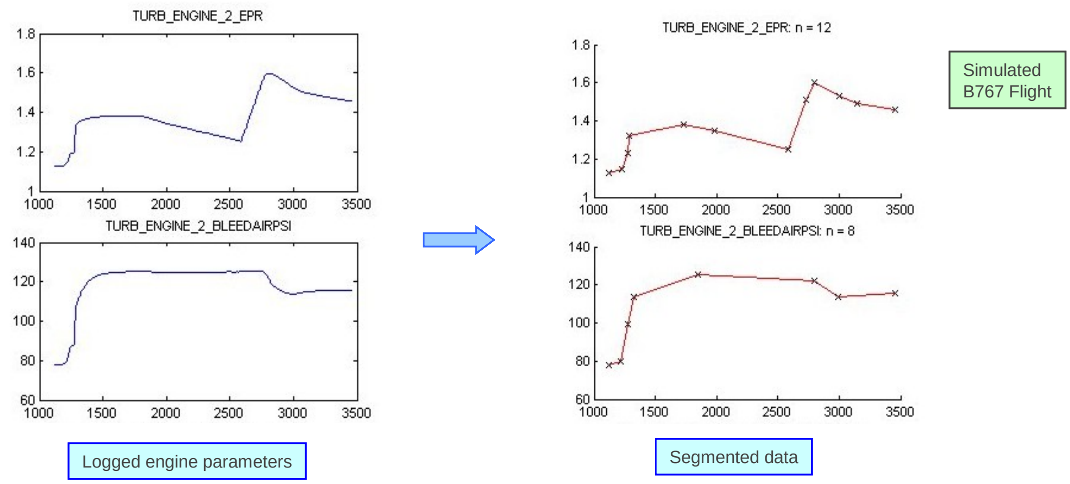


Figure 2: Examples of raw flight data channels with their segmented versions

The DAX symbol vocabulary begins with a horizontal segment. A horizontal segment is the simplest feature and is defined as a segment over which the magnitude of the slope is less than a user defined value. Adjacent horizontal features are combined into a single horizontal feature.

Level shift features can be increasing or decreasing. They are segments with a magnitude of slope exceeds a user defined value and the magnitude of the difference of signal values at the endpoints exceeds a second parameter. The combination ensures that level shifts are sharp and significant changes.

Spikes are simple to describe and difficult to parameterize. A spike begins with either an increasing or decreasing level shift. The level shift is followed immediately by another level shift of about the same magnitude in the opposite direction. Spikes are parameterized by slope, minimum amplitude percentage, maximum duration, and separation percentage. Two processing passes are used to ensure that a spike is not part of a train.

Trains are a pattern of spikes. CBA-OPS does not restrict how spikes are sequenced in a train. A train is simply defined as two or more spikes such that the extreme values between two consecutive spikes are separated by no more than a user defined distance along the time axis.

The remaining segments are either slowly increasing or decreasing. The magnitude of their slope is less than the minimum for a level shift and greater than the maximum for a horizontal segment. Segments with a positive slop are encoded as monotonic increasing. Those with a negative slope are monotonic decreasing segments.

DATA UNDERSTANDING

The Visualization Component provides an interactive visualization interface to help the user explore, identify, and visualize relevant aspects in flight data sets so s/he can formulate hypotheses. This is achieved through the following techniques and functionality elements:

- Visual encoding techniques associate visual representations (e.g., color, segment dimensions, modifiers) with data features of primary interest (e.g., extreme values, slope), facilitating the immediate recognition of feature properties in the flight data.
- Search capabilities help identify data segments containing specific features or combinations of features (e.g.,”find a rising segment longer than 4 seconds”).
- Data alignment and scaling support the synchronized analysis of events on different data channels, or on the same data channels on different flights. Specifically:
    - o Zooming allows data channels to be expanded as needed.
    - o Synchronized scrolling allows selected channels to scroll as one unit.
    - o Channel stacking allows channels of interest to be displayed next to each other for data comparison purposes.

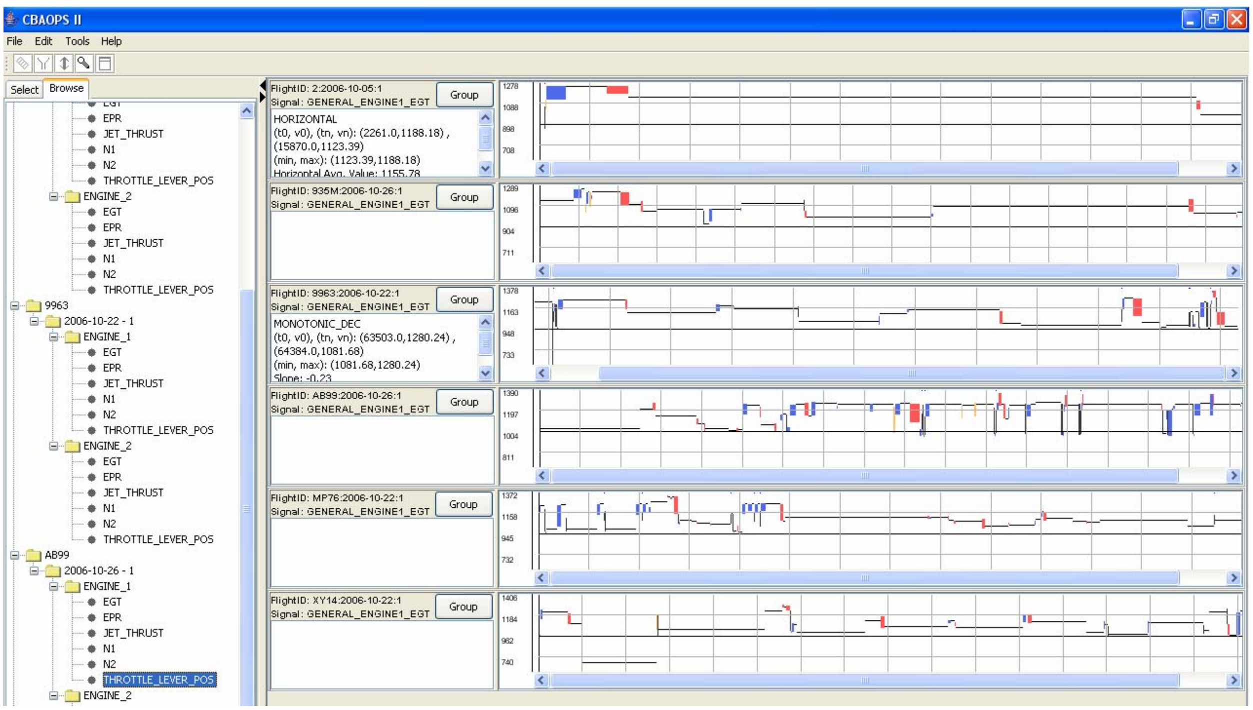


Figure 3: Visualization Component

The Visualization Component implements two signal visualizations and a browse capability for selecting signals to view. Figure 3 shows the browse tree on the left and a visualization based on the DAX symbol vocabulary. Each symbol in the DAX vocabulary is plotted as a graphic shape for easy recognition. Figure 4 illustrates the relationship between a time domain data plot of individual samples, a time domain plot of the segmented data, and the visualization of a DAX symbol train.

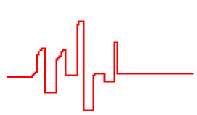
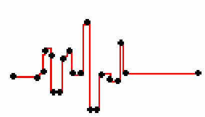
Raw View	Segmented View	Featurized View
		

Figure 4: Train Visualization

The time domain plot and the symbol visualization for a signal are compared in **Figure 5**. Blue and red boxes represent monotonic increasing and decreasing segments. The black boxes are level shifts. Lines at the top or bottom of the frame indicate increasing or decreasing. The amber box represents a train. Amplitude is show as the feature height and duration by the feature length. Relative values are indicated by feature placement on the vertical axis.

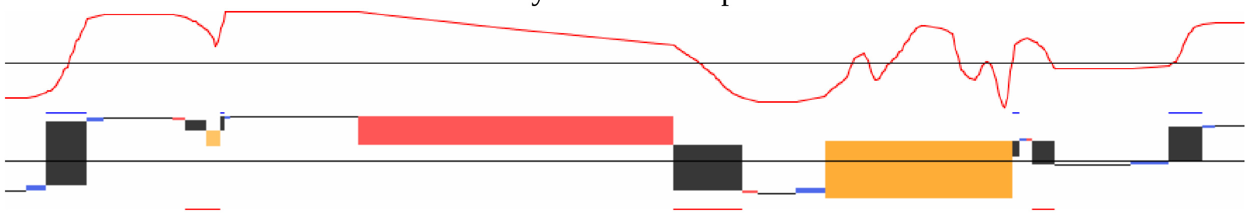


Figure 5: Visualization Comparison

The result is that the symbol visualization is a stylized version of the time domain trace that emphasizes the DAX symbols defined by domain experts as part of the ETL process.

DATA MINING

Constraint-based data-mining applies the apriori principle (Agrawal & Srikant, 1995) to ensure the computational tractability of data-mining operations. This principle says that any set of observations K that contains a frequent pattern P is a subset of K-1 that also contains P. Constraint-based data-mining uses a definition of P to begin data mining on K (Han et al., 1999).

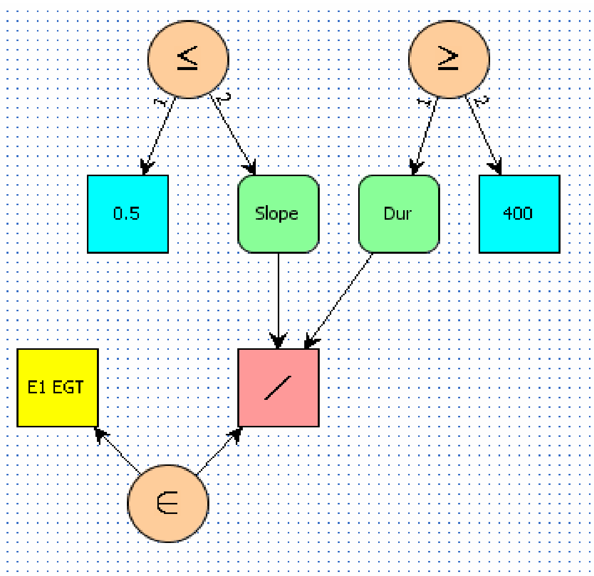


Figure 6: Hypothesis expressed in CDL

A Constraint Definition Language (CDL) was created to express expert knowledge in the form of hypotheses about patterns the characterize incidents of interest. The Constraint Definition Editor provides a graphical environment for defining hypotheses in CDL and adding constraints to them base on data-mining results.

CDL defines a constraint as a relation between one or more objects. Figure 6 shows a constraint that served as one hypothesis during system evaluation. The constraint specifies a level shift on the Engine 1 Exhaust Gas Temperature (EGT) sensor that has a slope of greater than ½ and duration of more than 50 seconds. (Sample rate = 8 Hz, 400 samples / 8 Hz = 50 seconds). Data-mining was performed to discover the additional relationships shown in Figure 7. A multivariate relation between the Engine 1 and Engine 2 EGT was discovered. The rate of temperature increase is less in engine 2 than in engine 1. This is shown in the lower section of the constraint graph. The upper group of nodes shows a temporal relationship that was also discovered through data-mining.

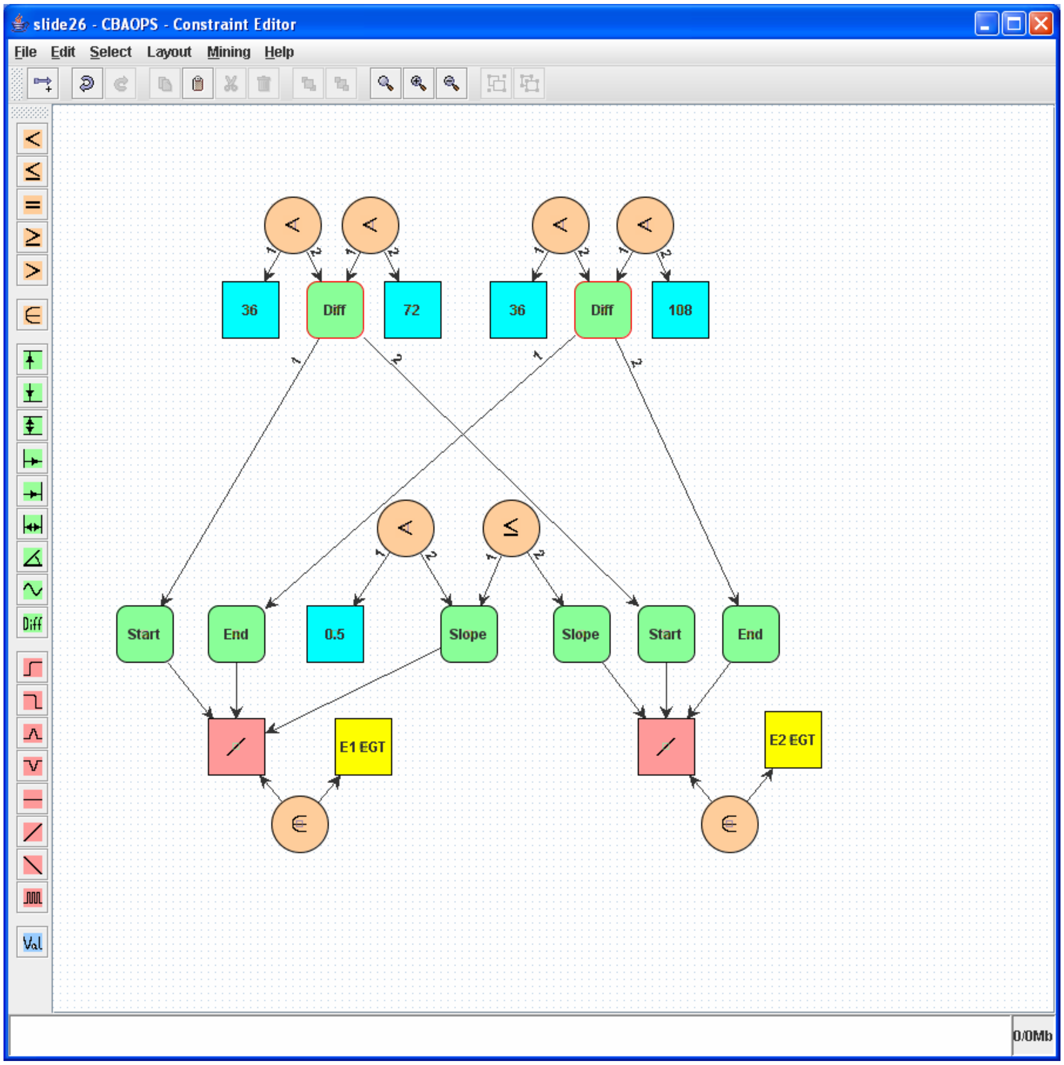


Figure 7: Constraint Editor showing hypothesis after data-mining

Figure 7 also shows the Constraint Editor. The purpose of the editor is to enable domain experts to formulate and edit constraint-based hypotheses for safety relevant flight conditions. The associated data-mining capabilities enable the expert to identify potential commonalities across data channels and complex relationships that characterize or are precursors to safety relevant flight conditions. The editor has the following components:

- A main menu bar to choose and save the hypothesis, channel, and other control operations.
- Several tool bars used to insert, edit, and link nodes

- The hypothesis work area (where the graph is formulated and edited)
- A status bar along the bottom that indicates the current action

The combination of the constraint editor and data-mining overcomes the challenges of acquiring and encoding large collections of expert knowledge. The exploratory nature of the tool encourages the domain expert express their knowledge in constraint graphs and to refine that knowledge through data-mining.

CONCLUSIONS

The system supports the detection and description of safety-relevant flight conditions in DFDR. The contributions of the approach are:

- A data reduction and transformation approach that scales down the high dimensionality of flight data sets to a size and format more easily explored by users and efficiently mined through algorithms.
- Visualization techniques that focus on relevant signal attributes, allow users to rapidly recognize potentially relevant signal areas, and develop initial hypotheses about safety-relevant flight conditions.
- A diagrammatic constraint editor that encodes hypotheses for incident description in a visually intuitive manner and in a format that can be used by data mining algorithms.
- Data mining components that support refining and expanding hypotheses into confirmed safety-relevant flight conditions by:
    - Identifying additional features that correlate with existing condition descriptions.
    - Uncovering additional relations between features in a description.
    - Enabling the refinement of a signal features' attribute values in a description.

This provides considerable flexibility by allowing the user to apply domain knowledge and expertise to guide the process of refining and expanding descriptions for safety-relevant flight conditions.

Future work is needed to increase the richness of the DAX process. The vocabulary of available time domain approximations and frequency domain approximations need investigation. The inclusion of data sources such flight operational and quality assurance (FOQA), logistics, or weather all offer the promise of fuller incident characterization. Model deployment also provides additional topics for investigation. A stream implementation of DAX will be required as will detectors for the discovered models. Investigation is needed to determine the feasibility and demonstrate the generation detector implementations such as decision trees, Bayesian belief networks (BBN), and dynamic belief networks (DBN).

ACKNOWLEDGEMENTS

This work was performed under Government contract number NNL04AA02C. This work was funded in its entirety by the NASA Langley Research Center's SBIR program. The authors would like to thank Ms. Rosa Oseguer-Lohr and Mr. Jon Jonsson for their support and eager engagement on this project. We would also like to acknowledge valuable feedback received from the Controls and Diagnostic Group at Pratt and Whitney.

REFERENCES

Agrawal, R., and Srikant, R. (1995) “Mining Sequential Patterns.” Proc. of 11th International Conference on Data Engineering.

Flight Safety Foundation (1998). “MD-83 Crew Continues Flight Operations with Suspected Engine Malfunction”, Accident Prevention, 55, 10, October 1998.

Han, Jiawei, Lakshmanan, Laks V. S., and Ng, Raymond T. (1999) “Constraint-based Multidimensional Data Mining”, IEEE Computer, August 1999.

Han, J, and M. Kamber (2000). Data Mining: Concepts and Techniques, Morgan Kaufmann.

Letourneau, S., Famili, F., Matwin, S. (1999) “Data Mining for Prediction of Aircraft Component Replacement.” IEEE Intelligent Systems Journal, December 1999

Lin, Jessica, Keogh, Eamonn, Lonardi, Stefano, and Chiu, Bill (2003) “A Symbolic Representation of Time Series, with Implications for Streaming Algorithms”, Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD03), June 13 2003.

Keller, Kirby, Wiegand, Dave, Swearingen, Kevin, Reisig, Chris, Black, Scott, Gillis, Alan, and Vandernoot, Mike (2001). “An Architecture to Implement Integrated Vehicle Health Management Systems.” Proceedings of AUTOTESTCON, 20-23 August 2001.

Malloy, D., Biegl, C., Zakrajsek, J., Meyer, C., and C. Fulton (1997). “Development of a Near Real-Time Turbine Engine Testing Diagnostic System Using Feature Extraction Algorithms,” XIII International Symposium on Air Breathing Engines, Chattanooga, TN.

Oats, Tim. (1999) “Identifying Distinctive Subsequences in Multivariate Time Series by Clustering”, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’99), August 1999.

Vlachos, Michail, Hadjieleftheriou, Marios, Gunopulos, Dimitrios, and Keogh, Eamonn (2003) “Indexing Multi-dimensional Time Series with Support for Multiple Distance Measures”, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’03), August 24-27 2003.

Williams, Zachary, “Benefits of IVHM: An Analytical Approach,” 1507, Proceedings of IEEE Aerospace Conference, 4-11 March 2006

Wilmering, Timothy J., and Ramesh, Anapathur V., “Assessing the Impact of Health Management Approaches on System Total cost of Ownership,” 1553, Proceedings of IEEE Aerospace Conference, 5-12 March 2005

Zakrajsek, J., Fulton, C., and C. Meyer (1994). “Feature Extraction for Post-Test Diagnostics,” Advanced Earth-to-Orbit Propulsion Technology Conference 1994, May