DEVELOPING DATA MINING-BASED PROGNOSTIC MODELS FOR CF-18

M. Zaluski, S. Létourneau, J. Bird, C. Yang

Proceedings of ASME Turbo Expo 2010, Glasgow, UK

Presented by Maria Pelevina

Machine Learning seminar, 17.06.2015, TU Darmstadt

Introduction

Generic methodology for prognostic models for complex systems

Developed in National Research Centre Canada since 1999

Applications

Aircrafts (1999)

Trains (2005)

Introduction

Generic methodology for prognostic models for complex systems

Developed in National Research Centre Canada since 1999

Applications

CF-18 – Fighting aircraft

Introduction – Motivation

Why CF-18?

Advanced diagnostic platform. In use since 1982.

1. Flight operational data
   - altitude, airspeed, engine temperature, vibrations, pressures, etc.
   - single diagnostic events
   - 5 years of records available as Aircraft Data Files
2. Maintenance data
   - what operations have been performed on which components
   - tracks over 2300 components
   - 10 years of records available as Oracle database

Research question

Can CF-18 available data be used to build prognostic models for different components?

Introduction – Task

What has been modeled?

Two components of an aircraft engine (model GE F404):

- No. 4 Bearing – reduces frictions between rotating parts
- Main Fuel Control (MFC) – controls fuel supply to combustion chambers

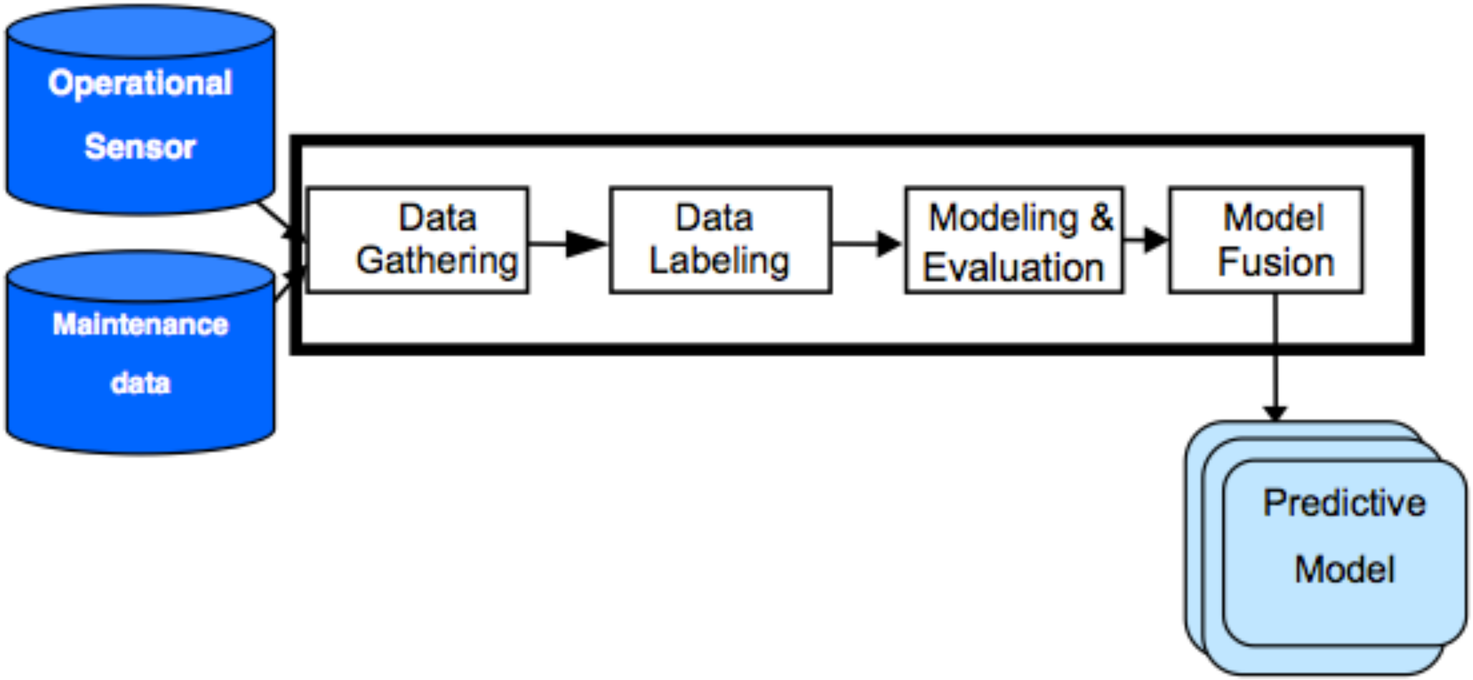
Problems

Delays, cancelations, engine damage, loos of engine or aircraft

Objective

- Enhance preventive maintenance procedures
- improve overall fleet readiness

Methodology – Overview



Methodology – Challenges

Data gathering

1. Too many measurements
   - advice from experts and reliable documentation
   - CF-18: 65 different message groupings in operations data (armament, engine performance, diagnostic information)

2. Time nature of instances
   - no random sampling
   - define failure event – CF-18: component replacement
   - extract instances in time window before and after event

3. Different recording frequencies
   - summarize records into single value
   - CF-18: speed → 1s, temperatures → 5s, pressure → 1s for 15s

Methodology - Challenges

Data labeling

1. Time nature of instances
   - time windows before and after the failure
   - optimal repair time, balance of negative/positive (at least 15% positive) are important

Modeling and evaluation

1. Simple, easy to explain models
   - simple algorithms: Naïve-Bayes, decision trees, rules

2. Time nature of instances
   - relevant for splitting in training/test datasets
   - relevant for evaluation

Model development – Data gathering

Maintenance data

Failure defined as a component replacement

Components

2300 components → 1. No. 4 B. 2. MFC

Replacements

1280 1700 → 42 6

Critical, replaceable

Same aircraft and engine for whole life cycle

Operational data

64 measurement sets → 4 measurement sets

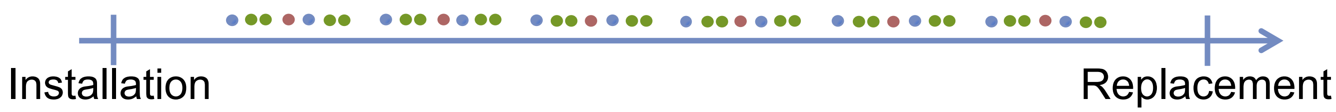
Relevant for engine performance

- 4 sets merged together
- time ordered
- filtered from flight number, date etc

Model development – Data gathering

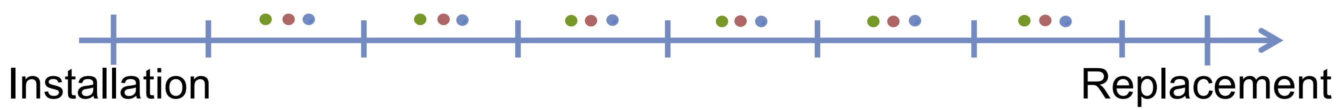
Maintenance data is used to extract relevant operational data.

1. Extract relevant records



- results in 42 Problem ID files for No.4 Bearing and 6 for MFC

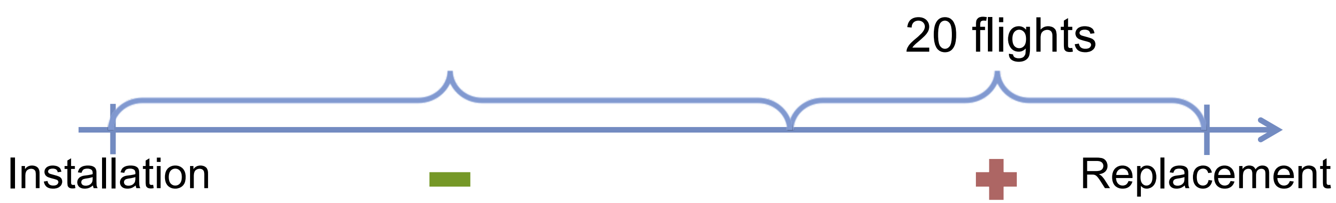
2. Group by flight and summarize



- 21 ways to summarize
  - traditional statistics (average, standard deviation)
  - ordered statistics (medians, quantiles)
- 21 datasets for each Problem ID file

Model development – Data labeling

3. Label the instances
   - 1 instance = summarized records of one flight



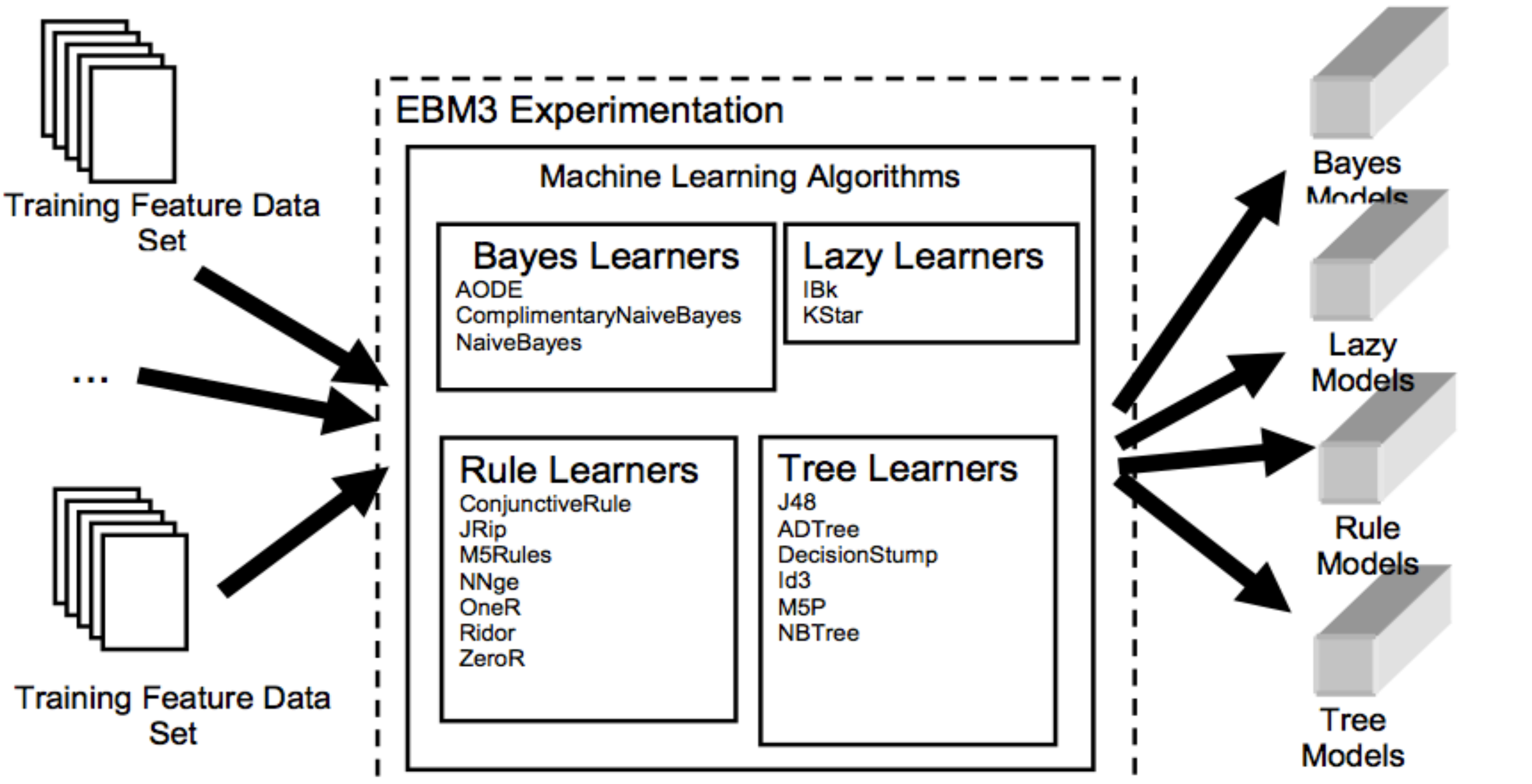
   - each Problem ID file sizes range from 32 to 343 instances

4. Combine and split the data

   For each component, for each of 21 summarization types:
   - combine all Problem ID series (ordered by ID)
   - split into train and test sets
     - No. 4 Bearing: 33 for training, 9 for testing
     - MFC: 5 for training, 1 for testing

Model development – Algorithms

Using WEKA implementation of algorithms.



Model development – Automation

EBM3 – Environment for Building Models for Machinery Maintenance

Each step of methodology defined as independent component

- input/output formats, parameters, algorithms
- written in various languages (Java, Perl, R, Python, C/C++)

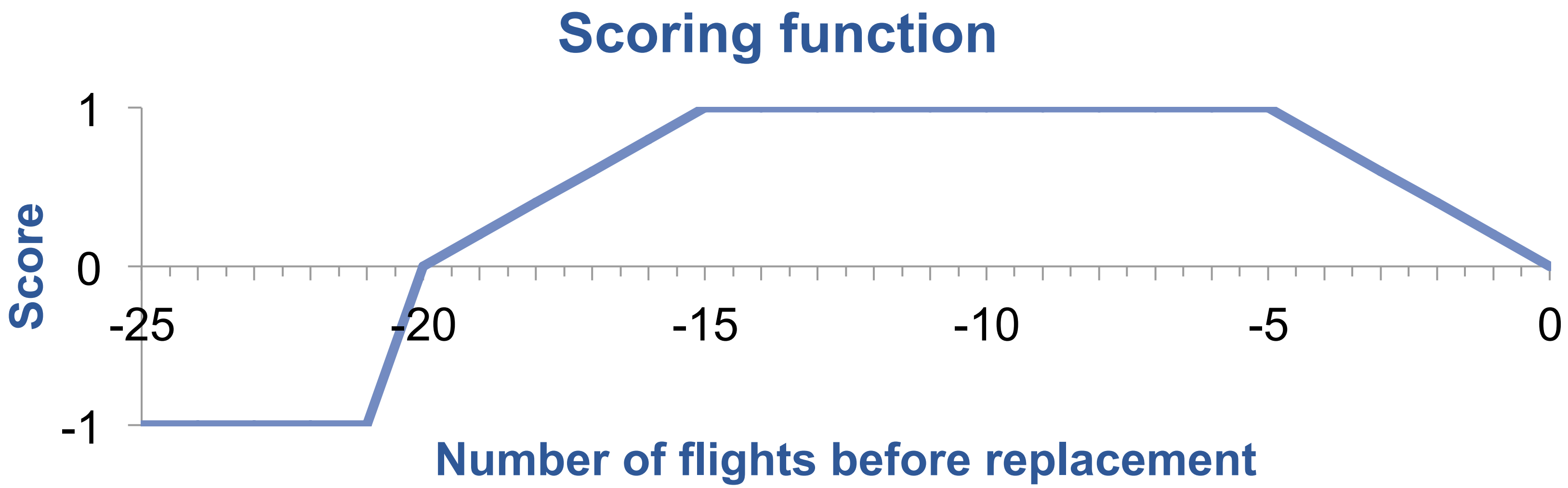
EBM3:

1. launches components, connects computers if necessary
2. converts data formats between components
3. handles iterations and stores results for exploration
   - various parameters, datasets and learning algorithms
4. deploys desired model (software and configuration) to end users

Evaluation – Criteria

Score

- sum of scores for each positive flight classification



Problem detection rate

- proportion of “detected problems”
- problem detected, if at least one of the flight in Alert Period is classified as positive

Evaluation – Criteria

The Paper

Accuracy

$$\cancel{\text{Precision}} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

$$\text{Recall} = \frac{\begin{array}{c}\text{Number of correct} \\ \text{replacement alert predictions}\end{array}}{\text{Total number of replacement alerts}}$$

Earlier Paper

Accuracy

$$\cancel{\text{Precision}} = \frac{\text{Number of correct predictions}}{\text{Total number of instances}}$$

true positive

$$\text{Recall} = \frac{\begin{array}{c}\text{Number of correct} \\ \text{positive predictions}\end{array}}{\text{Total number of positive instances}}$$

true positive +
false negatives

Evaluation – Results

negative/positive instances: 50/50 (20 first and 20 last flights)

Feature Data Set	Algorithm	Performance Measures				
		Score max. 126	Problem Detection Rate	Accuracy ~~Precision~~	Recall	
Interquantile Average 10 90	Conjunctive Rule	-4	33%	51%	8%	
	NNge	33	100%	61%	72%	
	NBTree	49	100%	61%	68%	fpr: 23%
Upper Quantile Average 75	Conjunctive Rule	-200	100%	49%	87%	
	NNge	73	100%	59%	44%	fpr: 13%
	NBTree	-18	100%	56%	58%	

No. 4 Bearing

Problems get detected, but not precisely in the Alert Period

Evaluation – Results

negative/positive instances: 85/15

Leave One Batch Out cross validation

Feature Data Set	Algorithms	Performance Measures			
		Score	Problem Detection Rate	Precisio n	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%

fpr: 26%

Main Fuel Control

Evaluation – Discussion

1. No best model for all evaluation criteria
   - MFC – not enough data
   - MFC – imbalance

2. Much space for improvement
   - custom metrics instead of precision and recall
     - alert precision, latency, assessment of false alerts
   - feature evaluation
     - using domain knowledge or WEKA
   - feature engineering
     - transform raw measurements into more relevant parameters using physics knowledge

Related Work

1999 [1]

- failures of Airbus 320 components (16 unidentified components)
- no process automation, IR for maintenance data
- no feature transformation, no model fusion,
- similar results: no one best model (score evaluation only)

2005 [2]

- failures of train wheels
- no process automation, no feature transformation
- model fusion – stacking
- results: meta model has 97% detection rate, 8% fpr, highest score

2014 [3]

- predict energy consumption for air conditioning systems in buildings
- continuous values – regression tree, SVM
- results: incomparable

Conclusion

Negative

1. Results not promising
   - no economical value, if more false alerts, than truthful predictions
   - no cost evaluation
2. Questionable development decision
   - only 20 first and 20 last flights for No. 4 Bearing
   - metrics not informative enough

Positive

1. Summarization of typical challenges
2. Automation of experiments

Question

- Reasons for low results: methodology or data?

Thank you for your attention!

Questions?

Bibliography

[1] Sylvain Létourneau, Fazel Famili, and Stan Matwin.
Data Mining to Predict Aircraft Component Replacement.
IEEE Intelligent Systems 14, 6 (November 1999), 59-66.

[2] Chunsheng Yang and Sylvain Létourneau.
Learning to predict train wheel failures.
In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (KDD '05). ACM, New York, NY, USA, 516-525.

[3] Chunsheng Yang, Sylvain Létourneau, Hongyu Guo.
Developing Data-driven Models to Predict BEMS Energy Consumption for Demand Response Systems.
IEA/AIE (1) 2014: 188-197