Managing Streamed Sensor Data for Mobile Equipment Prognostics

Authors: Toby Griffiths, Dr Debora Correa, A/Prof Adriano Polpo and Prof Melinda Hodkiewicz

Presenters: Dr Debora Correa and Prof Melinda Hodkiewicz

Intelligent Maintenance Conference, 8 & 9th September, Zurich
Project Motivation

- 125M connected cars on our roads by 2022\(^1\)
- Streamed sensor data on mining trucks and excavators now
- Equipment manufacturers’ predictive maintenance solutions immature
- Asset operators
  - Raw data difficult to access and process
  - Are not getting **data-driven** insights for maintenance actions

---

\(^1\) CounterPoint Research [1] Image source: Tesla [2], Komatsu [3]
Data Sources

- Dataset of 13 excavators measured over 9 months totalling ~300M rows in 19.9GB
- 58 numeric sensors and 40 binary indicators (per excavator)
- Fleet management database
- Operator entered machine status codes
- CMMS work orders describing maintenance events

Image source: Hitachi [4]
Raw Data State – normal transformation

### Long form sensor data

<table>
<thead>
<tr>
<th>Time</th>
<th>Sensor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:33:24</td>
<td>57</td>
<td>4.176</td>
</tr>
<tr>
<td>0:33:24</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>0:33:24+</td>
<td>58</td>
<td>30537.4</td>
</tr>
<tr>
<td>0:33:24</td>
<td>3</td>
<td>59</td>
</tr>
<tr>
<td>0:33:54</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>0:33:54+</td>
<td>58</td>
<td>30537.4</td>
</tr>
<tr>
<td>0:33:55</td>
<td>3</td>
<td>59</td>
</tr>
<tr>
<td>0:33:55</td>
<td>57</td>
<td>4.128</td>
</tr>
<tr>
<td>0:34:24</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>0:34:24</td>
<td>58</td>
<td>30537.4</td>
</tr>
<tr>
<td>0:34:24</td>
<td>57</td>
<td>3.792</td>
</tr>
<tr>
<td>0:34:25</td>
<td>3</td>
<td>60</td>
</tr>
</tbody>
</table>

An example of raw data for three sensors with varying timestamps.

### Pivot sensor data

<table>
<thead>
<tr>
<th>Time</th>
<th>Sensor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td>T+0.5</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>T+0.3</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>T+1.3</td>
<td>2</td>
<td>B</td>
</tr>
</tbody>
</table>

### Fleet management data

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T+1.3</td>
<td>Loading</td>
</tr>
<tr>
<td>T+1.3</td>
<td>T+2.7</td>
<td>Hydraulic fault</td>
</tr>
</tbody>
</table>

### CMMS data

<table>
<thead>
<tr>
<th>Start Time</th>
<th>Work order short text</th>
</tr>
</thead>
<tbody>
<tr>
<td>T +12.2</td>
<td>Repair hydraulic system</td>
</tr>
</tbody>
</table>

### Wide form data (for data analysis)

<table>
<thead>
<tr>
<th>Time</th>
<th>Sensor 1</th>
<th>Sensor 2</th>
<th>Loading</th>
<th>Hydraulic fault</th>
<th>CMMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>X</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T+0.3</td>
<td>NA</td>
<td>A</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T+0.5</td>
<td>Y</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T+1.3</td>
<td>NA</td>
<td>B</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>T+12.2</td>
<td>Z</td>
<td>C</td>
<td>NA</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Size of the Data

- This example spans **four days**, showing data for one sensor on one excavator.
- The complete dataset encompasses 58 numeric sensors and 40 binary indicators across 272 days for 13 excavators.
- This extract can be considered **0.001%** of the data.
- Impossible for engineers to monitor this volume of data visually.
<table>
<thead>
<tr>
<th>Failure signature?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hydraulic Oil Tank Temperature</strong></td>
</tr>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Note:** The images represent various graphs and diagrams related to hydraulic oil systems, indicating different parameters such as temperature, contamination, overheat, and oil level alarms.
How can streamed sensor data be efficiently transformed and compressed and then used to inform maintenance on mobile equipment?
Data Cleaning Pipeline

**Transform**
- Extract timestamps and values to separate files
- Pivot each separately

**Compress**
- OHLC (Open High Low Close) data representation

**Model**
- Engineering based dimension reduction
- Log-linear mean regression model

Data per sensor + CMMS data + Fleet Management data

Clean and separate into 146 files, one for each sensor, alarm, fleet management status and CMMS data

30 MB as opposed to 17GB
OHLC Data Transformation

- Original data has 14402 entries.
- OHLC data has 600 entries.
- Retains key trends of the original data.
Human in the loop

Sensor Dashboard

I think there is a hydraulic fault...

Add hydraulic fault code to the Fleet Management system

There is a hydraulic fault, please create notification...

Work Order

approved

Create order notification

<}>

systemhealthlab.com
Response Variable

Both response variables are generated by an operator

Hydraulic Fault from the Fleet Manager System

Maintenance work order for hydraulic system repair (CMMS)

Comparison of two response variables
Modelling event occurrence

Fleet Management

CMMS
Time to event modelling

- Developed a predictive survival model using time varying covariates
- Log linear regression model
What did we learn from the modelling?

- Traditional models for event prediction were overfitted and had poor predictive power
- For our log linear regression model the most influential covariates for predicting hydraulic faults based on response variables from a) fleet management and b) CMMS data are **not**
  - Hydraulic Oil Tank Temperatures
  - Pilot Pump Pressure
  - Hydraulic Oil Level Low Alarm
  - Hydraulic Oil Overheat Indicator
  - Pump Contamination Indicator

- This is less surprising than you might first think. There is no pattern in the data as the response variable is about a human (the operator) perception about an event that has not happened yet.
Conclusions

• Messy data – most of our time was devoted to transforming and compressing the dataset
• If you don’t have an actual failure, what are you using as your response variable? Is a human involved?
• Both fleet management and CMMS data are based on operators:
  – perceptions (what they consider potential failures), and
  – actions (when they report them)
• Human data are not consistent for prediction purpose
• Data analysts may not realise that the fleet management entry is human generated
• Lots of time is being wasted in trying to build predictive models on asset data sets with poorly defined response variables
• Asset owners should consider having Data Engineers to ensure consistent pre-processing of their data and selection of ’good’ response variables
Thank you!

Contact: Prof Melinda Hodkiewicz
melinda.hodkiewicz@uwa.edu.au

Dr. Débora Correa
debora.correa@uwa.edu.au

Acknowledgements:
Our industry partner and the CTMTDS
BHP Fellowship for Engineering for Remote Operations