Processing of Unstructured Short Text in Maintenance Records

NIST Standards Requirements Workshop for Natural Language Analysis
May 2018

Yiyang Gao, Caitlin Woods, Melinda Hodkiewicz, Tim French, Wei Liu
University of Western Australia – BHP Fellowship funded Siri for Maintenance Project
ARC ITTC for Transforming Maintenance through Data Science
Motivation

- Thousands of maintenance tasks are performed a year in most asset intensive industries.
- What was done to assets including when, where, why and how is captured in short unstructured text.
- The knowledge in this text is vital to improving maintenance efficiency and asset reliability.

Image from: https://www.oceanasolutions.org/the-importance-of-mechanical-maintenance/
Problem

- Time Consuming
- Inconsistent
- Domain knowledge required

Can this process be semi-automated?

Unstructured maintenance records  Structured machine readable maintenance records
Example of maintenance record text

Technicians Perform Maintenance Task

Technicians Fill in Comment

Reliability Engineer Perform Analysis

Short Text

• replace t/con pump
• positioner arm out of alignment
• repair broken lift cylinders

Action, Item

Item, Item symptom

Action, Item, Item state
A pipeline for semi-automatic processing of unstructured short text into a structured format
Experimental Setup

- Approx 700,000 historical maintenance records
- Domain → Maintenance of Heavy Mobile Equipment

<table>
<thead>
<tr>
<th>WOType</th>
<th>PONumber</th>
<th>BscStartDate</th>
<th>ActStartDate</th>
<th>ShortText</th>
<th>WorkCentre</th>
<th>FuncLoc</th>
<th>FuncLocDesc</th>
<th>UserStatus</th>
<th>SystemStatus</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM01</td>
<td>1</td>
<td>10/07/2005</td>
<td>10/07/2005</td>
<td>replace lube tank fill hose.</td>
<td>WC1</td>
<td>FL1</td>
<td>grease lube system</td>
<td>Removed</td>
<td>Removed</td>
<td>651</td>
</tr>
<tr>
<td>PM01</td>
<td>2</td>
<td>10/07/2005</td>
<td>10/07/2005</td>
<td>checked rotary head alignment</td>
<td>WC1</td>
<td>FL2</td>
<td>mast</td>
<td>Removed</td>
<td>Removed</td>
<td>434</td>
</tr>
<tr>
<td>PM01</td>
<td>3</td>
<td>10/07/2005</td>
<td>10/07/2005</td>
<td>reweld shock sub to rotary head</td>
<td>WC1</td>
<td>FL3</td>
<td>subs/adaptors</td>
<td>Removed</td>
<td>Removed</td>
<td>361</td>
</tr>
<tr>
<td>PM01</td>
<td>4</td>
<td>27/07/2005</td>
<td>28/07/2005</td>
<td>centre tank not filling with water</td>
<td>WC1</td>
<td>FL4</td>
<td>water injection system</td>
<td>Removed</td>
<td>Removed</td>
<td>144</td>
</tr>
<tr>
<td>PM01</td>
<td>5</td>
<td>2/07/2001</td>
<td>2/07/2001</td>
<td>hydraulic pump failure</td>
<td>WC2</td>
<td>FL5</td>
<td>implement pump</td>
<td>Removed</td>
<td>Removed</td>
<td>NA</td>
</tr>
<tr>
<td>PM01</td>
<td>6</td>
<td>1/07/2001</td>
<td>2/07/2001</td>
<td>pump drive box coupling and shaft u/s.</td>
<td>WC2</td>
<td>FL6</td>
<td>#NAME?</td>
<td>Removed</td>
<td>Removed</td>
<td>172</td>
</tr>
<tr>
<td>PM01</td>
<td>7</td>
<td>11/07/2005</td>
<td>0/1/00</td>
<td>repair engine excessive blowby</td>
<td>WC1</td>
<td>FL7</td>
<td>engine</td>
<td>Removed</td>
<td>Removed</td>
<td>0</td>
</tr>
<tr>
<td>PM01</td>
<td>8</td>
<td>16/07/2005</td>
<td>0/1/00</td>
<td>two broken growser bolts r/h side</td>
<td>WC1</td>
<td>FL8</td>
<td>track - right</td>
<td>Removed</td>
<td>Removed</td>
<td>19</td>
</tr>
<tr>
<td>PM01</td>
<td>9</td>
<td>27/07/2005</td>
<td>0/1/00</td>
<td>adjust all head slides</td>
<td>WC1</td>
<td>FL9</td>
<td>rotary guide frame</td>
<td>Removed</td>
<td>Removed</td>
<td>0</td>
</tr>
<tr>
<td>PM01</td>
<td>10</td>
<td>27/07/2005</td>
<td>28/07/2005</td>
<td>change out main gearbox oil to 85/140</td>
<td>WC1</td>
<td>FL10</td>
<td>oil - pump drive gearbox</td>
<td>Removed</td>
<td>Removed</td>
<td>72</td>
</tr>
<tr>
<td>PM01</td>
<td>11</td>
<td>27/07/2005</td>
<td>28/07/2005</td>
<td>change out both final drive oil to 85/140</td>
<td>WC1</td>
<td>FL11</td>
<td>final drive assemblies</td>
<td>Removed</td>
<td>Removed</td>
<td>144</td>
</tr>
<tr>
<td>PM01</td>
<td>12</td>
<td>27/07/2005</td>
<td>28/07/2005</td>
<td>change out rotary head gearbox oil</td>
<td>WC1</td>
<td>FL12</td>
<td>oil - rotary gearbox</td>
<td>Removed</td>
<td>Removed</td>
<td>72</td>
</tr>
</tbody>
</table>

Data of interest: short text in maintenance records
Proposed Pipeline

1. Text Pre-processing Module
   - Pre-processed Corpus
     - Item Symptom/State Detection Module
       - Condition Tagged Corpus
         - Maintenance Item Detection Module
           - Item, Condition Tagged Corpus
             - Maintenance Activity Detection Module
               - Structured Maintenance Records
Text Pre-processing Module

Unstructured Maintenance Records

Tokenization

Semantic Transformation

Normalization & Abbreviation Disambiguation

Lemmatization

Pre-processed Maintenance Records

'trd2034 replacx broken t/converters wont work 08/12/2015'

['trd2034', 'replacx', 'broken', 't/converters', 'wont', 'work', '08/12/2015']

['equipment_id', 'replacx', 'broken', 't_converters', 'cannot', 'work', 'date_time']

['equipment_id', 'replace', 'broken', 'torque', 'converters', 'cannot', 'work', 'date_time']

['equipment_id', 'replace', 'broken', 'torque', 'converter', 'cannot', 'work', 'date_time']

Pre-processed Maintenance Records
## Text Pre-processing Result

<table>
<thead>
<tr>
<th>Task</th>
<th>Size of Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokenization</td>
<td>54288</td>
</tr>
<tr>
<td>Semantic Transformation</td>
<td>20200</td>
</tr>
<tr>
<td>Normalization &amp; Abbreviation</td>
<td>12743</td>
</tr>
<tr>
<td>Disambiguation</td>
<td></td>
</tr>
<tr>
<td>Lemmatization</td>
<td>11618</td>
</tr>
</tbody>
</table>
Symptom/State Detection Module

Pre-processed Maintenance Records

Symptom/State Detection Module

Bigram Detection

Keyword based bigram filtering

Symptom/State Dictionary Building

Tagging

What is bigram for symptom/state detection?

[equipment_id', 'replace', 'broken', 'torque', 'converter', 'cannot', 'work', 'date_time']

[equipment_id', 'replace']

[replace', 'broken']

[broken', 'torque']

[torque', 'converter']

[pump', 'is', 'broken']

[pump', 'is']

[is', 'broken']

[converter', 'cannot']

[cannot', 'work']

[work', 'date_time']

[cylinders', 'not', 'lifting']

[cylinders', 'not']

[not', 'lifting']

Symptom/State Tagged Maintenance Records
Symptom/State Detection Module

Pre-processed Maintenance Records

Symptom/State Detection Module

- Bigram Detection
- Keyword based bigram filtering
- Symptom/State Dictionary Building
- Tagging

Head Word Filtering
- Cannot <XXXX>
- Not <XXXX>
- Is <XXXX>
- Are <XXXX>
- Be <XXXX>

Cannot start
Cannot release
Cannot raise
Cannot work
Cannot open

Symptom/State Tagged Maintenance Records
Maintenance Item Detection Module

- Symptom/State Tagged Maintenance Records

**Maintenance Item Detection Module**

- Bigram Detection
- bigram filtering
- Tagging

Bigram that does not contain

1. Symptom/State,
2. Maintenance activity,
3. Common stop words

Is a bigram for maintenance item

- ['equipment_id', 'replace', 'broken', 'torque', 'converter', 'cannot', 'work', 'date_time']
- ['equipment_id', 'replace']
- ['replace', 'broken']
- ['broken', 'torque']
- ['torque', 'converter']
- ['converter', 'cannot']
- ['cannot', 'work']
- ['work', 'date_time']

Symptom/State, Item Tagged Maintenance Records
Maintenance Activity Detection Module

Symptom/State Tagged Maintenance Records

Verb Conjugation

Look up & Tag

Pre-defined List of Maintenance Verb

Symptom/State, Item, Activity Tagged Maintenance Records

replaced → replace
replacement → replace
Unstructured Maintenance Records

Text Pre-processing Module

Pre-processed Corpus

POS Tagging

POS Tagged Corpus

Item, Activity, Symptom/State detection

Rule Based Detection
1. If POS is ‘VB’, → Maintenance activity
2. If POS is ‘ADJ’, → Item symptom/State
3. If POS is ‘N’, → Maintenance item

Average Perceptron Tagger Trained on Brown corpus

Structured Maintenance Records
Performance Evaluation

Validation Corpus
- Size: 360
- Human Processed
  - Normalization
  - Abbreviation Disambiguation
  - Item symptom/state tagging (#)
  - Maintenance Item tagging (~)
  - Maintenance activity tagging (=)

Jaccard Index for similarity measurement

\[
J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]

A: Expert tagged
B: Pipeline tagged

Raw: ‘replacx broken arm rest’

A: [‘replace=’, ‘broken#’, ‘arm~rest’]

B: [‘replace=’, ‘arm~rest’]

\[
\text{Number of Common Tokens} \quad \frac{\text{Total Number of Unique Tokens}}{}
\]

Jaccard Index = \(\frac{2}{3} = 66\%\)
## Result

<table>
<thead>
<tr>
<th>Task</th>
<th>Proposed Pipeline Result</th>
<th>Base Pipeline Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Symptom/State Detection</td>
<td>63.2%</td>
<td>8.9%</td>
</tr>
<tr>
<td>Maintenance Item Detection</td>
<td>54.9%</td>
<td>14.5%</td>
</tr>
<tr>
<td>Maintenance Activity Detection</td>
<td>88.9%</td>
<td>24.4%</td>
</tr>
<tr>
<td>Overall</td>
<td>56.7%</td>
<td>12.8%</td>
</tr>
</tbody>
</table>

Proposed Pipeline > Base Pipeline
Next steps

Our next steps

• Working on improving the code
• Explore LSTM and other methods
• Work on other corpus
• Building on this work through our new $8.8M industry and government funded Centre for Transforming Maintenance through Data Science
  https://www.maintenance.org.au/

Proposed next steps for this community

• Develop shared test sets
• Develop shared dictionaries specifically for maintenance including lists of synonyms, misspellings, etc.
• Share code, papers etc. (e.g. on GitHub)
• Assist each other in peer reviewing code