Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021 - ISBN – 978-1-936263-34-9

Semi-automated Estimation of Reliability Measures from Maintenance Work Order Records

Tyler Bikaun and Melinda Hodkiewicz

The University of Western Australia, Perth, Western Australia, 6009, Australia
tyler.bikaun@research.uwa.edu.au
melinda.hodkiewicz@uwa.edu.au

Abstract

Determining mean-time-between-failure (MTBF) estimation for in-service assets is an essential process. Statistical distributions of end-of-life values are used to assess asset reliability performance and the effectiveness of maintenance strategy. However, identifying the end-of-life event for each instance of functional failure is an arduous, manual process dependent on structured and unstructured fields in the maintenance management system and rules used by individual reliability engineers. We emulate the process of end-of-life event detection using a natural language processing pipeline followed by statistical parameter estimation to produce MTBF values for in-service assets from maintenance work order data. Using this pipeline we test how alternate mappings of words in unstructured text and the use, or not, of structured data can impact the identification of end-of-life events. We demonstrate the pipeline on data sets from two industrial users with 14,508 and 89,259 maintenance work orders, respectively. We find that resulting MTBF estimates can vary from, for example, 97 to 226 days for a single asset depending on the mappings and rules used. The main contributions of this paper are a) a demonstration of the impact of undocumented decisions made by reliability engineers in identifying end-of-life events on MTBF estimates, and b) an end to end pipeline for MTBF estimation from raw maintenance work order texts. In addition we provide open-source code[1] for the pipeline that can be used by asset owning organisations to semi-automate the MTBF estimation process in a manner that is fast and scalable, and ensures the rules used for end of life determination are documented and hence the process is transparent and repeatable.

1. Introduction

"You can't manage what you can't measure" is often attributed to W. Edwards Deming and as a concept underpins organizational performance management systems. For asset intensive organisations, a common measure for monitoring the performance of a physical asset or asset class is reliability. Reliability is usually expressed as the probability that an asset or component will perform its intended function for a specified time period under specified conditions (Department of Defence, 1981). Two common metrics for reliability are mean time to failure (MTTF) and mean time between failure (MTBF), for non-repairable and repairable assets respectively (Meeker & Escobar, 2014). There are a number of methods for determining MTTF and MTBF. A simple method for estimating mean life is using a point estimate calculated by dividing the operating time to failure by the number of assets run to failure (SMRP, 2017). Matters get more complicated when equipment is repairable, data is truncated, and there are suspensions as well as failures in the data. However even the point estimate method can be challenging to calculate on in-service assets given how asset data are recorded particularly how an end-of-life event is identified. Determining the time from start-of to end-of-life for each instance of functional failure is an arduous, manual process dependent on structured and unstructured data fields and the expertise of the reliability engineer(s). A consequence of this situation is that many asset intensive organisations do not have trusted, regularly updated values for the MTTF/MTBF of their assets (Hodkiewicz & Ho, 2016; Lukens, Naik, Saetia, & Hu, 2019). This is metaphorical equivalent of flying blind for reliability managers.

This paper is organised as follows. Section 2 presents a review of relevant literature. Section 3 describes details of two real-world industrial data sets. Section 4 presents the proposed pipeline and implementation details. Section 5 details experiments performed. Section 6 presents results of the pipeline. Section 7 discusses and evaluates the pipeline. Lastly, conclusions are presented in Section 8.

[1] https://code-ittc.csiro.au/tyler.bikaun/mtbf_from_mwo

2. LITERATURE REVIEW

To overcome the challenges faced by subject matter experts when determining MTTF/MTBF estimates for in-service assets, semi- and fully-automatic pipelines built using natural language processing (NLP) have been proposed (Lukens et al., 2019). These pipelines typically employ rule-based expert systems or data-driven annotation methods to structure information contained within the unstructured fields of maintenance records to enable reasoning over maintenance activities and observations, as well as supporting event type characterisation (Hodkiewicz & Ho, 2016; Sexton, Hodkiewicz, Brundage, & Smoker, 2018).

Rule-based expert systems consist of sets of hand-crafted rules developed by subject matter experts to elicit knowledge from maintenance records. These systems have proven to be effective for structuring maintenance records to support reliability parameter estimation (Hodkiewicz & Ho, 2016) as well as identifying high-level ideological concepts in unstructured free-text fields (Gao, Woods, Liu, French, & Hodkiewicz, 2020). However, the construct of these systems is resource intensive and rigid making it challenging to adapt across domains such as between organisations and industries, rendering them ill-suited as a general class of techniques.

Instead of relying on rigid rule-sets to structure unstructured fields for down-stream tasks, statistical and machine learning (ML) based approaches have been used. As a result, rules are learnt explicitly, as features, from data directly (Sexton et al., 2018). However, a caveat of data-driven approaches is their dependency on large amounts of high-quality annotated data that is scarce in industrial maintenance settings. Regardless, ML approaches have been used similarly to rule-based systems with varying levels of success to support extraction of concepts for root cause analysis (Navinchandran, Sharp, Brundage, & Sexton, 2019), event classification (Arif-Uz-Zaman, Cholette, Li, Ma, & Karim, 2016; Arif-Uz-Zaman, Cholette, Ma, & Karim, 2017), and reliability parameter estimation (Sexton et al., 2018). The success of these techniques has largely depended on the availability of data and the representation method used in the ML algorithms i.e. how unstructured texts are represented numerically. Moreover, due to the unavailability of annotated data that is representative of industrial maintenance as a whole, similar to rule-based systems, ML approaches are yet to be proven as a domain-adaptable technique in industrial maintenance due to their inability to widely generalise.

To learn effectively, the manner that data is represented for ML algorithms is vital. Historically, representation of unstructured fields in maintenance records have largely been count or co-occurrence based such as bag-of-words (Arif-Uz-Zaman et al., 2016, 2017) or numerical features derived from latent semantic analysis (Sharp, Sexton, & Brundage, 2016). More recently, embedding and language models have been adopted to transfer learn representations learnt from common English domains to domain-specific applications such as event type classification (Khabiri, Gifford, Vinzamuri, Patel, & Mazzoleni, 2019; Cadavid, Grabot, Lamouri, Pellerin, & Fortin, 2020) and degradation modelling (Yang, Baraldi, & Zio, 2020) in industrial settings. Although, the use of such techniques remains inconclusive as to whether the representations used non-industrial settings are widely effective in domain-specific tasks due to the unavailability of large annotated data sets.

In this work, we propose a pipeline using the well established word embedding technique word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) coupled with flexible domain-specific rules to emulate the process a typical reliability engineer would perform when estimating MTTF/MTBF. Our pipeline is easily domain-adaptable and requires very little computational and subject matter expert resources. Using this pipeline, we are able to investigate the impact of decisions made when processing maintenance records on resulting MTTF/MTBF estimates, and is the first study of this kind to our knowledge.

3. DATA SETS DESCRIPTION

There are a number of time-stamped sources for asset life information, for example, the computerised maintenance management system (CMMS), downtime accounting system (DAS) (also called a process historian), fleet management system (FMS), maintenance logs and warranty records, as well as ad hoc tables stored in databases and spreadsheets by engineers and maintainers recording events on particular equipment of interest. These database systems record events, usually unwanted events, either directly in the case of DAS, FMS and warranty reports or in the case of CMMS the record is a notification of the need for work in response to an event or asset condition. CMMS data is usually entered and managed by maintenance personnel, DAS and FMS by operating personnel, and warranty records by dealers.

Our focus in this study is on leveraging data captured in the CMMS system for MTTF/MTBF estimation as this is the data source commonly used by reliability engineers responsible for assets in process plants such as found in the chemical and mining industries, power and water treatment plants (Hodkiewicz & Ho, 2016; Lukens, Naik, Hu, Doan, & Abado, 2017). CMMS data is time stamped and describes transactional events such as actual and desired work activities, parts orders, payments, resourcing, and costs.

Historical MWO data for pump assets were acquired from CMMS of two large mining and mineral processing plants operating in Western Australia[2]. The data sets consist of 14,508 and 89,259 work orders for plants A and B, respectively, with each consisting of both *structured* and *unstruc-*

[2]These plants will be referred to as *A* and *B* throughout.

tured data fields that are extracted to support statistical life data analysis. An overview of each data set is provided in Table 1.

To conserve lexical diversity, only light pre-processing is performed on the unstructured data such as removing casing and non-alphanumerical characters (except special characters - and /). Special characters are kept as they are heavily used in both data sets to represent compound words and abbreviations. For example, *de-contactor*, *v-belt*, *a/c* (*air conditioner*), and *s/w* (*south west*). Tokenization of each document is performed by splitting on white space. For structured data such as resource and time estimates, only date-time objects are normalised.

Table 1. Overview of historical MWO datasets.

Plant	Asset Type	MWOs	Mean Tokens
A	Pump	14,508	5.5
B	Pump	89,259	8.0

3.1. Structured data fields

Structured fields are typically auto-generated by CMMS systems based on predefined or calculated values. Fields pertaining to i) work order classification (e.g. preventative, breakdown, corrective), ii) event dates (e.g. basic actual start date of maintenance activity), iii) asset identification and description, iv) actual craft hours, and v) actual total cost, are extracted to support contextual filtering of work orders using domain logic.

3.2. Unstructured data fields

The unstructured descriptive short text field used by maintenance personnel to capture in-field observations provides a rich body of contextual information about maintenance activities performed on in-service assets (Gao et al., 2020; Brundage, Sexton, Hodkiewicz, Dima, & Lukens, 2020). A unique property of these texts is their ability to be structured with natural language processing such that verbs and adjectives corresponding to end-of-life events in MWOs can be identified and used as evidence for statistical life data analysis.

4. Pipeline

The proposed pipeline uses NLP and domain logic to emulate the typical workflow of a reliability engineer when determining MTBF estimates[3]. The use of NLP and domain logic provides a consistent and standardised foundation to programmatically perform sensitivity analysis to gain insight into decisions that affect MTBF estimations.

Our model consists of two stages, firstly an unsupervised embedding model is trained on MWOs to learn domain-specific word associations that are used to construct an entity dictionary (gazetteer) that maps EOL terms to words in the unstructured field of MWOs. Using the domain-specific gazetteer in conjunction with domain logic, eligible MWOs are classified as either failures or suspensions to support reliability parameter estimation using 2-parameter Weibull analysis. An overview of the model is shown in Figure 1.

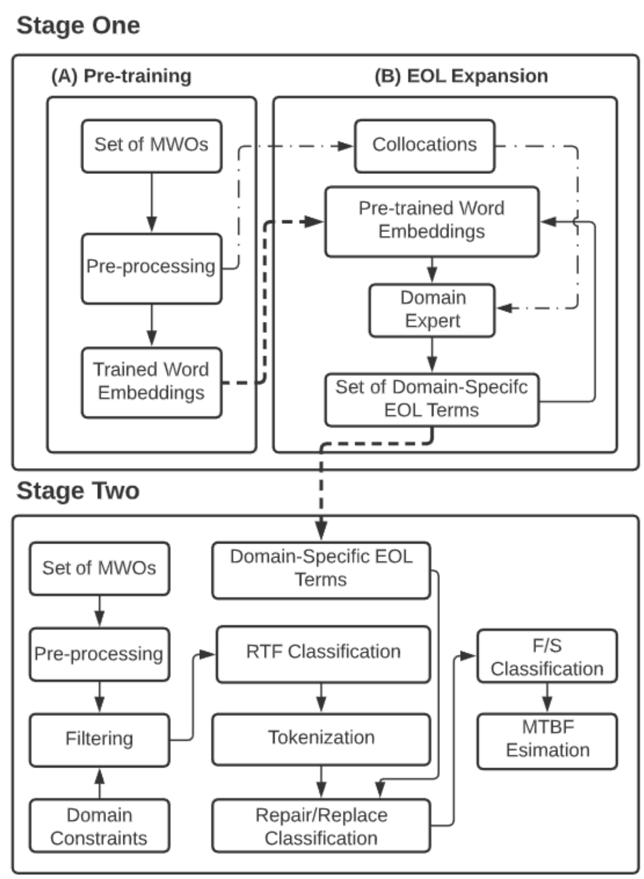

Figure 1. Overview of proposed pipeline (RTF refers to *return-to-function*, and F/S refers to *failure or suspension*).

Table 2. Examples of gazetteer developed in stage 1 of the pipeline.

Term	Action
replcmt	replace
u/s	replace
snaped	replace
change out	replace
not pumping	replace
repair	repair
remount	repair

4.1. Stage One: Gazetteer Construction using Word Embedding

A gazetteer is a classical technique in the field of information extraction that acts as a conceptual dictionary that maps words to concepts or entities. This technique has shown to

[3]We focus on MTBF as the assets under study are repairable, however the pipeline is applicable for non-repairable assets.

Table 3. Worked example of filtering and classification steps using a cost threshold of $2,000 (TAC refers to *total actual cost*).

Description	**Order Type**	**TAC** ($)	**RTF**	**EOL**	**Materialised**	**F/S**
$ B11 1A2B Mech Seal leaking steam on pu	Corrective	20,359	Reactive	True	True	Failure
C8 Piping design for Sulphur flow meters	Corrective	8,300	Reactive	False	True	-
3M Mech Lube Pump 1A2B	Preventative	310	Proactive	False	False	-
C9 JDI 1A2B not pumping	Corrective	1,339	Reactive	True	False	-
B11 1A2B Sulphur leaking from seal	Breakdown	30,832	Reactive	True	True	Failure
B11 Replace 1A2B Sulphur pump MJ	Corrective	36,237	Reactive	True	True	Failure
B12 1A2B Vibration Check / Housekeeping	Corrective	39,809	Reactive	False	True	-
B12 1A2B Pump not performing	Corrective	27,545	Reactive	True	True	Failure
$ C9 Replace pump 1A2B	Breakdown	12,466	Reactive	True	True	Failure

be effective in identifying concepts in unstructured maintenance records such as item, activity, and state (Gao et al., 2020), however the construction was resource intensive due to the need for large rule-sets that can be challenging to update when concepts shift. In this work, gazetteer construction using word embeddings is performed in two steps.

Firstly, a word2vec (Mikolov et al., 2013) word embedding model is trained using self-supervision on all available unstructured MWO text. This allows word associations to be learnt under the distribution hypothesis i.e. "you shall know a word by the company that it keeps" (Harris, 1954). Learnt word associations allow words that frequently co-occur to be numerically close to one another. Consider the following three work orders: *b11 a2b1 - repair pump*, *b11 repl pump repr leaks 1a22b*, and *b16 a2b-21b replace pump*, if given enough similar examples, a word embedding model would deduce that the words *replace*, *repl*, and *repair* are very similar as they all frequently co-occur with *pump*. If trained on an entire history of MWOs, an embedding model would learn these associations between all unique words, including those with lexical variations such as incorrect spelling and abbreviations. Word embedding has become a popular way of numerically representing texts in industrial settings (Khabiri et al., 2019; Cadavid et al., 2020; Yang et al., 2020), however to our knowledge they have not been used as a domain-adaptable way of constructing gazetteers in this setting. We favour word2vec due to its representative power, computational efficiency and ability to capture lexical variations and abbreviations that may be troublesome for techniques that have fixed vocabularies (Khabiri et al., 2019).

Secondly, term expansion is performed using the pre-trained embedding model to populate the gazetteer that is used to map EOL terms to words in unstructured MWO text. The construction process starts with a single EOL seed term such as *replace* and recursively finds similar words using the embedding model that are validated and added to the gazetteer by a domain-expert through the use of a command-line interface, that continues until the terms are exhausted. For example, the term expansion process starting with *replace* could take the form of: $replace \rightarrow \{u/s, rpl, ...\}$, u/s[4] $\rightarrow$ $\{u/s, rpl, failed, leaking, ...\}$, and so forth. In addition to this, pertinent n-gram EOL terms are identified by the domain-expert from a list of high-frequency collocations derived from the set of unstructured texts. Collocations are sets of words that frequently co-occur, for example, *not* and *pumping* in *not pumping*, and *not* and *performing* in *not performing*. This process aims to emulate the terms that a reliability engineer would use to contextually filter work order descriptions in spreadsheet software. Examples of terms captured by this process are exhibited within Table 2.

[4] u/s refers to unserviceable

4.2. Stage Two: Statistical Life Data Analysis

With domain-specific EOL terms identified, evidence for statistical life data analysis are captured by emulating the process of a typical reliability engineer. This process consists of 4 filtering and classification steps, each aiming to ask a question about the work order:

1. Does the unstructured field indicate an EOL event? → use gazetteer of domain-specific EOL terms to determine if verbs or adjectives constitute a replace or repair event.
2. Is it likely to have materialised to an EOL event, and is it associated to the functional or physical component? → filter on minimum total actual cost or total actual craft time.
3. Is the work proactive or reactive? → classify return-to-function (RTF) based on structured work order classification field.
4. Is it a failure or suspension event? → classify as failure (if EOL event and corrective), suspension (if EOL event and proactive), or other (if neither).

To illustrate this process, we provide a worked example in Table 3 showing the result of the steps applied to 9 real work orders. After identification of suitable evidence, and verification that the i.i.d assumption holds, parameter estimation using 2-parameter Weibull analysis including censoring is performed and MTBF estimations are obtained.

5. EXPERIMENTS

Experiments are performed on the data sets specified in Table 1 which consist of pump assets used in the mining and minerals industry. For each experiment we filter MWOs with actual total cost and/or actual total craft time as $2,000 and 8 hours, respectively. We treat cost preferentially as it is more indicative of material change whereas hours are not, for example, an inspection in some instances could take much longer than a replacement activity, whereas it's unlikely an inspection will cost more than a replacement event when duration is similar. For plant A, both structured fields were available so cost is used, whereas for plant B only time information was available. Furthermore, the minimum threshold of failure and/or suspension events required for an individual asset is set at five [5]. In the code accompanying this work, all of these parameters can be modified and are use case specific.

Additionally, to gain insight into the impact of decisions made by reliability engineers when determining MTBF estimates, we experiment with three different, albeit extreme, scenarios that are defined as:

- S1: gazetteer with only *replace* term,
- S2: expanded EOL gazetteer, and
- S3: expanded EOL gazetteer and structured data fields.

Each scenario is performed on the three most populous asset types within each data set (Table 4).

Table 4. Exemplar assets used in sensitivity analysis.

Plant A	Plant B
A1 - Centrifugal pump	B1 - Warman 8/6 FAH
A2 - Piston pump	B2 - Worthington 10LR15A
A3 - Peristaltic pump	B3 - Warman 10/6 FM

6. Results

6.1. Pipeline performance

The pipeline was applied to both industrial data sets having 14,508 and 89,259 work orders, respectively. This resulted in MTBF estimates for 93 and 669 pumps as shown in Table 5 from a set of 903 and 3079 pumps. Construction and validation of the gazetteer using word embeddings and collocations took less than 10 minutes for both data sets, and the pipeline was able to process MWOs for each plant in 22 and 91 seconds, respectively. This compares to manual processing by a reliability engineer that would take weeks for this number of pumps.

6.2. Variation in MTBF estimates

The results of the three experiments, which mimic different decisions made by a reliability engineer when processing MWO data, are shown in Table 6. It is clear that decisions made in the processing of MWOs for statistical life analysis can result in substantial variations to MTBF estimations. This is seen in data set A where MTBF vary from 97 to 226 days for a single asset depending on whether the engineer selected pipeline S1, S2 or S3. The impact is less marked on data set B as the rules used in the B1 and B2 process resulted in data sets with only suspensions and no failures. This is a reflection of the use of fixed interval replacement strategy used at the site of the B pump set.

Table 5. Overall results of the pipeline applied to each data set.

	Plant A	**Plant B**
Assets	903	3079
Samples	3112	35170
Failures	1874	6746
Suspensions	14	2850
Eligible Assets[6]	93	669
Compute Time	22s	91s

Table 6. Impact of decisions made on MTBF estimates for exemplar pump assets (T, F and S refer to *total*, *failure*, and *suspension*, respectively).

Plant	**Scenario**	**Asset**	**MWOs**	**MTBF** (days)
A	S1	A1	58T/15F/0S	184.8
	S2	A1	58T/29F/0S	97.1
	S3	A1	20T/9F/0S	226.2
	S1	A2	90T/17F/0S	181.9
	S2	A2	90T/30T/0S	101.9
	S3	A2	40T/14F/0S	209.9
	S1	A3	31T/11F/0S	244.3
	S2	A3	31T/15F/0S	191.7
	S3	A3	10T/7F/0S	408.5
B	S1	B1	134T/0F/23S	-
	S2	B1	134T/0F/23S	-
	S3	B1	115T/9F/16S	185.6
	S1	B2	-	-
	S2	B2	10T/7F/0S	318.8
	S3	B2	15T/6F/2S	334.1
	S1	B3	-	-
	S2	B3	137T/1F/16S	-
	S3	B3	114T/5F/11S	301.4

6.3. Examination of identified EOL events

Table 7 provides a list of all the work orders identified by the three experiments for a centrifugal sulphur pump (A1). The results considered the most closely aligned to the processing of an experienced reliability engineer are shown in the right hand column and are the outputs of pipeline S3.

There are 13 MWOs in this S3 column. On review by experts most of these are plausibly related to a functional failure of the pump with the exception of two MWOs for *delete replace motor/gearbox on 1a2b* and *b16 repl motor switchgear g/box 1a-2b*. In addition to these two MWO's that were incorrectly

[5] This corresponds to the lower limit specified by Natrella *et al.* (2010) for there to be statistical significance.

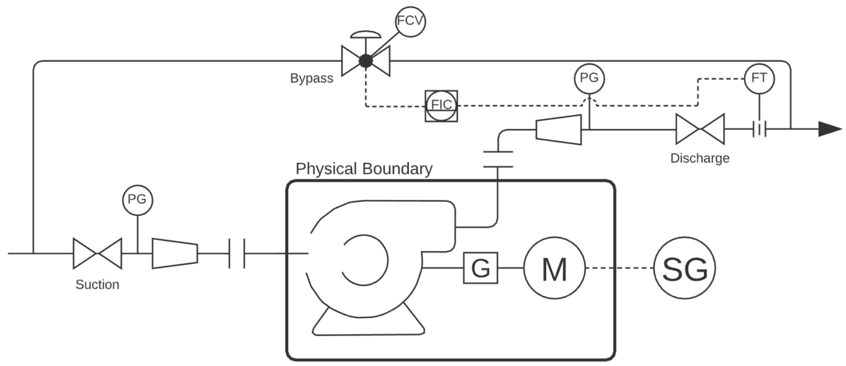

Figure 2. Simplified P&ID of A1 that highlights the boundary between physical and functional systems (G - gearbox, M - electric motor, SB - switch board).

Table 7. Work orders classified as failures for A1 under each scenario (coloured entries indicate those that do not exist in S3).

S1 (*replace* only)	**S2 (term expansion)**	**S3 (term expansion and structured fields)**
b11 replace 1a2b sulphur pump mj	b11 replace 1a2b sulphur pump mj	b11 replace 1a2b sulphur pump mj
c9 1a2b replace worn/ corroded bolts	1a-2b is not pumping	1a-2b is not pumping
jds event replace spring on pump	1a2b not pumping	1a2b not pumping
c8 1a2b replace drain valve lc-0123	delete replace motor/gearbox on 1a2b	delete replace motor/gearbox on 1a2b
b12 replace bypass spring 1a2b	b11 1a2b mech seal leaking steam on pu	b11 1a2b mech seal leaking steam on pu
b12 replace bypass spring 1a2b	b11 1a-2b not pumping v/well	b11 1a-2b not pumping v/well
fabricate spool to replace nrv at h2s	b11 1a2b sulphur leaking from seal	b11 1a2b sulphur leaking from seal
b12 ab-0010 and ab-0020 replace valve	b11 steam leaks 1a-2b	b11 steam leaks 1a-2b
c9 replace pump 1a2b	b12 1a-2b pump not pumping	b12 1a-2b pump not pumping
replace pump 1a2b	b16 repl motor switchgear g/box 1a-2b	b16 repl motor switchgear g/box 1a-2b
discharge pressure guage replace	c9 replace pump 1a2b	c9 replace pump 1a2b
cub prjcts vsd replacement tie ins	pump is not pumping	pump is not pumping
delete replace pump 1a2b	replace pump 1a2b	replace pump 1a2b
remove replace flanges	c9 jds 1a2b not pumping	
delete replace motor/gearbox on 1a2b	c9 1a2b replace worn/ corroded bolts	
	jds event replace spring on pump	
	c8 1a2b replace drain valve ab-0120	
	b12 replace bypass spring 1a2b	
	b16 split suction line flanges 1a/1b	
	b12 replace bypass spring 1a2b	
	fabricate spool to replace nrv at h2s	
	1a2b not pumping	
	b12 ab-0010 and ab-0020 replace valve	
	b11 investigate not pumping issues 1a2b	
	ab-0120 holed steam jacketed line/valve	
	discharge pressure guage replace	
	cub prjcts vsd replacement tie ins	
	delete replace pump 1a2b	
	remove replace flanges	

assigned as EOL events, we also examined all the outputs of the S3 pipeline for A1 to check if there were failures or suspensions that might have been missed. These are shown in Table 8, one can see that they are largely related to operational events and we know from the rules used in the S3 pipeline that none cost more than $2,000 or 8 hours and so feel comfortable that they were excluded from the analysis. Should a reliability engineer feel that these represented genuine EOL events then it is possible to change the rules used in the pipeline to ensure they are included.

There are only 4 MWOs in common between S1 and S3, and 11 in S1 that do not occur in S3. These 11 entries are in S1 column are not valid EOL events for the pump. They relate to ancillary components associated with the pump that are functional, but do not contribute to life of the primary asset. These MWO's are identified by the S1 pipeline because the MWO contains the word *replace* and the use of a functional location code to capture all work associated with the functional system. This notion of function versus physical system is illustrated in Figure 2. It shows that all components listed are considered part of A1. Our interest here is in only

MWO's associated with the reliability of the pump such as EOL events associated with the impeller, casing, coupling, mechanical seal, bearings and motor. We want to exclude MWOs associated with the switchgear, pressure gauge, drain valve and bypass system.

The S2 pipeline expands the terms related to EOL events in that we see an expansion of work orders (15 to 29) due to the breadth of terms that can be mapped to MWOs. Newly acquired terms capture more phenomenological concepts as well as lexical variants, for example *leaking*, *repl*, and *holed*. Examples include *holed steam jacketed line/valve* and *h16 split suction line flanges 1a/1b*. In the latter the term *split* has been interpreted by the computer as an adjective instead of a verb.

However because no rules are used, the resulting S2 list contains many items that are not considered EOL events as they are not associated with minimum cost and/or time hurdles to count as significant events. Examples include the MWO *c9 jds 1a2b not pumping* that while almost identical to one considered eligible *1a-2b is not pumping* cost only $1,340 that is below the rule used of $2,000. Upon further inspection, the cost associated with this MWO only consisted of labour, not materials. This indicates the MWO may have data quality issues or the issue was resolved without material change. In contrast, the eligible MWO has a cost of $12,517 of that $10,704 is for materials and $1,813 for labour.

6.4. Pipeline evaluation

To better understand how the pipeline discriminates between eligible and non-eligible MWOs, we randomly sampled historical MWOs from 20 assets in each data set for evaluation after running them through the S3 pipeline. Subject matter experts then assessed each MWO to determine if they described an eligible failure or suspension and how it corresponded to the classification provided by the pipeline. The resulting F_1 scores[7] for data set A and B were 79.3% and 54.3%, respectively. We use F_1 score as our evaluation metric as the number of samples in each data set are unbalanced, for data sets A and B, the distribution of eligible and non-eligible MWOs are 70% / 30% and 36% / 64%, respectively.

Table 8. Description of work orders not classified as either failure or suspension for A1 in S3.

Description
b12 1a2b vibration check / housekeeping
b11 1a2b pump output restricted
c8 piping design for sulphur flow meters
b16 sce rotable spare rv assy 1a2b
b12 1a2b poor performace
b12 1a2b pump not performing
b11 1a2b pump trips on high amps

[7] F_1 is the harmonic mean of precision and recall e.g. $F_1 = 2 \times \dfrac{P \times R}{P + R}$

Compared to similar classification tasks performed on maintenance records (Sharp et al., 2016; Cadavid et al., 2020), our semi-automated pipeline performed better than expected. Although, the evaluation identified two shortcomings of the pipeline that can be attributed to a) an inability to strongly differentiate between functional and physical components, and b) a reliance on structured cost information. These shortcomings are particularly marked for data set B due to the unavailability of cost information, instead craft hours were used to reason over MWOs, resulting in impinged efficacy of the pipeline. This result is due to craft hours not being as strong an indicator as cost for identifying manifestations of EOL events. For example, consider the record *b21 1a2b pump not performing*, if this was accompanied with either 8 craft hours or an actual total cost of $25,000, the craft hours would not be strongly indicative of an EOL event as they could be attributed to various activities such as a change out or an inspection. On the contrary, the total actual cost would be more strongly indicative of material change of the pump that gives confidence that an EOL event has occurred.

7. Discussion

Decisions made when processing MWOs for statistical life analysis can have considerable impact on the resulting MTBF estimates for pumps in our industry supplied data sets, as demonstrated in Table 6. The absence of a clearly defined ground truth for determining an EOL event from these MWO records is a key contributor to this situation. In practice, to assess an EOL event, reliability engineers needs to communicate with the maintainer(s) working on the pumps to ascertain the extent of the damage and details of work done to restore the function. This may be possible in small facilities with a limited number of assets but is not possible in the facilities that contributed data for this study with 903 and 3070 pumps as well as other equipment. The only information available to the reliability engineer are these MWO records, possibly delay accounting records, external repair shop records and, in the case of serious consequent events, a root cause analysis. However, bringing all of these data sets together is a time-consuming manual process and so the normal state of affairs is to look at the MWO records as we have shown.

We show the MTBF estimates resulting from different processing pipelines to identify EOL events. Pipeline S1 uses simple keywords such as *replace*, S2 uses a gazetteer populated through contextual term expansion, and S3 uses a gazetteer and structured fields. There is a 233% variation between the MTBF values resulting from S2 and S3 for pump A1, with S2 suggesting an MTBF of 97 days and S3 of 226 days. This demonstrates how sensitive estimations are in the presence, or not, of structured fields. There are also significant differences the use of simple terms as in the S1 pipeline producing an MTBF estimate of 244 days for pump A3 compared to the S3 methods resulting in an estimate of 408 days. In practice, the implications of actioning MTBF values that are either over, or under, estimated will be detrimental to the effectiveness of a site's maintenance strategy. For example, over-estimation of MTBF estimates can result in unexpected downtime as preventative maintenance schedules will be less frequent and unwanted corrective events may occur resulting in a drain of people and resources from scheduled work. On the contrary, if MTBF estimates are too low, then maintenance budgets may become over capitalised, due to an increased need for maintainers and parts to perform unnecessary work.

There has been much push back in the maintenance community against using data stored in CMMS for reliability estimates due to perceptions of poor quality (Hodkiewicz & Ho, 2016). Our work shows that data necessary to identify potential EOL events can be identified using the unstructured text. However, for the potential EOL event to be confirmed additional fortuitous data must be available. Specifically we have relied on cost as being informative of whether the work done was significant or not. When the data is fit for purpose (data set A), our pipeline exhibits acceptable performance (79.3% F_1). We suggest this is close to that of a reliability engineer. However, to the best of our knowledge, human-level performance for eliciting evidence from MWOs for MTBF estimation remains unstudied and unknown. For plant B the cost data was unavailable resulting in the total craft time being used to qualify identified EOL events. However, it proved unable to reasonably discriminate between functional and physical components in unstructured texts. The impact of this is shown by the 25% difference in F_1 score between both data sets.

One advantage of our pipeline is that it does not rely on detailed annotation of unstructured fields in MWO like supervised concept extraction methods (Sharp et al., 2016; Brundage, Morris, Sexton, Moccozet, & Hoffman, 2018; Sexton et al., 2018), or on subject matter experts for cluster alignment such as in unsupervised clustering (Salo, McMillan, & Connor, 2019), it easily scales to large amounts of MWOs whilst requiring minor computational and subject matter expert resources. This is highlighted by the end-to-end execution time of our pipeline (pre-processing, term expansion and parameter estimation) only requiring 6 and 10 minutes for data set A and B, respectively. The end-to-end application speed of our pipeline shows a marked difference compared to aforementioned methods, as both require substantial resource requirements for eliciting annotations or correcting predictions. Moreover, as the gazetteer used in our pipeline is constructed from scratch using an embedding model trained directly from domain-specific MWOs, our pipeline is easily domain adaptable across organisations, industries, and CMMS systems. However, our pipeline does not constitute a general model for these settings, it solely provides a convenient and resource efficient means of application between different settings that in contrast to other methods still poses a challenge.

Lastly, we release the code for our pipeline to the community with the aim of it being added to practitioners metaphorical toolbox and to push the community towards open-source, reproducible research. Our code is written in Python using open-source packages, with verification of our statistical life data analysis process including censoring events through bench marking against popular commercially available reliability software. The code is also configurable to different CMMS systems and business constraints through a common configuration file.

8. Conclusion

In this paper, we demonstrate the sensitivity of MTTF/MTBF estimates to decisions made by reliability engineers when processing pump data held in maintenance management systems. We use NLP and domain-specific rules to emulate these decisions and note that these decisions are often undocumented and hidden to potential users of the resulting reliability metrics. Our work suggests there should be greater transparency and consistency in how reliability estimates are determined in individual organisations otherwise large variations in estimations can be produced. We release the code accompanying this work to support reliability engineers with a fast, scalable, transparent and configurable means of obtaining MTTF/MTBF estimations from fields commonly found in maintenance work order data. Future work will focus on expanding the applicability of this model to more asset types as well as removing the subject matter expert from the gazetteer construction process entirely.

Acknowledgment

Both authors acknowledge funding from the Australian Research Council Centre for Transforming Maintenance through Data Science (Industrial Transformation Research Program Grant No. IC180100030). Additionally, Bikaun acknowledges funding from the Mineral Research Institute of Western Australia, and Hodkiewicz acknowledges funding from the BHP Fellowship for Engineering for Remote Operations.

References

Arif-Uz-Zaman, K., Cholette, M. E., Li, F., Ma, L., & Karim, A. (2016). A data fusion approach of multiple maintenance data sources for real-world reliability modelling. In *Proceedings of the 10th world congress on engineering asset management (wceam 2015)* (pp. 69–77).

Arif-Uz-Zaman, K., Cholette, M. E., Ma, L., & Karim, A. (2017). Extracting failure time data from industrial maintenance records using text mining. *Advanced Engineering Informatics*, *33*, 388-396.

Brundage, M. P., Morris, K., Sexton, T., Moccozet, S., & Hoffman, M. (2018). Developing maintenance key performance indicators from maintenance work order data. In *International manufacturing science and engineering conference* (Vol. 51371, p. V003T02A027).

Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2020). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*.

Cadavid, J. P. U., Grabot, B., Lamouri, S., Pellerin, R., & Fortin, A. (2020). Valuing free-form text data from maintenance logs through transfer learning with camembert. *Enterprise Information Systems*, 1–29.

Department of Defence. (1981). *Definitions of terms for reliability and maintainability* [Standard]. Washington, DC: Department of Defense.

Gao, Y., Woods, C., Liu, W., French, T., & Hodkiewicz, M. (2020). Pipeline for machine reading of unstructured maintenance work order records. In *Proceedings of the 30th european safety and reliability conference and the 15th probablistic safety assessment and management conference.*

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162.

Hodkiewicz, M., & Ho, M. T.-W. (2016). Cleaning historical maintenance work order data for reliability analysis. *Journal of Quality in Maintenance Engineering*.

Khabiri, E., Gifford, W. M., Vinzamuri, B., Patel, D., & Mazzoleni, P. (2019). Industry specific word embedding and its application in log classification. In *Proceedings of the 28th acm international conference on information and knowledge management* (pp. 2713–2721).

Lukens, S., Naik, M., Hu, X., Doan, D. S., & Abado, S. (2017). The role of transactional data in prognostics and health management work processes. In *Proceedings of the annual conference of the prognostics and health management society* (pp. 517–528).

Lukens, S., Naik, M., Saetia, K., & Hu, X. (2019). Best practices framework for improving maintenance data quality to enable asset performance analytics. In *Annual conference of the phm society* (Vol. 11).

Meeker, W. Q., & Escobar, L. A. (2014). *Statistical methods for reliability data*. John Wiley & Sons.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*, 3111–3119.

Navinchandran, M., Sharp, M. E., Brundage, M. P., & Sexton, T. B. (2019). Studies to predict maintenance time duration and important factors from maintenance workorder data. In *Proceedings of the annual conference of the phm society* (Vol. 11).

Salo, E., McMillan, D., & Connor, R. (2019). Work orders - value from structureless text in the era of digitisation. *SPE Offshore Europe*.

Sexton, T., Hodkiewicz, M., Brundage, M. P., & Smoker, T. (2018). Benchmarking for keyword extraction methodologies in maintenance work orders. In *Phm society conference* (Vol. 10).

Sharp, M. E., Sexton, T. B., & Brundage, M. P. (2016). Semi-autonomous labeling of unstructured maintenance log data for diagnostic root cause analysis.

SMRP. (2017). *Society of Maintenance and Reliability Professionals (SMRP) Best Practices 5th. Edition* (Vol. 5; Standard). Atlanta, GA: Society for Maintenance and Reliability Professionals.

Yang, Z., Baraldi, P., & Zio, E. (2020). A novel method for maintenance record clustering and its application to a case study of maintenance optimization. *Reliability Engineering & System Safety*, *203*, 107103.

BIOGRAPHIES

Tyler Bikaun lives in Perth, Western Australia and is a Ph.D. candidate at The University of Western Australia. He received his bachelors degree (honours) in mechanical engineering from Curtin University in 2018 where he focused on numerical modelling of dynamic systems. Prior to commencing post graduate studies, he worked for Wood PLC as an engineering consultant in the resources sector. Tyler's primary research interest is in technical language processing for low-resource domains with an emphasis on maintenance in mining and mineral processing sectors.

Professor Melinda Hodkiewicz Melinda Hodkiewicz ATSE graduated with a BA(Hons) in Metallurgy and Science of Materials from Oxford University in 1985 and a Ph.D. in Engineering from the University of Western Australia (UWA) in 2004. Her early career was spent in operations and maintenance roles in the mining industry in the USA, Australia and the UK. She has been the BHP Fellow for Engineering for Remote Operations at the UWA since 2015 and has an ongoing teaching and research position at UWA. She leads Theme 1 program on "Support the Maintainer" in the A$8.8M Government and Industry-funded Centre for Tranforming Maintenance through Data Science. She is a Chartered Engineer registered in the UK, a member of IEEE and the IOM. She sits on the Advisory Board for Australia's National Offshore Petroleum Safety Authority and METS Ignited, the Australian Government's National Growth Centre for Mining, Equipment and Technology. In 2012-2015 she led Australia's effort to develop the ISO 55000-55002 set of Asset Management standards. She is a member of the editorial boards for IEEE Access, IMechE Journal of Risk and Reliability, and the Journal of Quality in Maintenance Engineer. She has held visiting fellowships at NIST in the US and the Turing Institute in London. In 2016 she was awarded the MESA Medal, a lifetime achievement award for services to Asset Management in Australia and in 2019 she was made a Fellow of the Australian Academy of Technology and Engineering.