Technical Language Processing of Long Text Maintenance Records

Brad Hershowitz[1]

Melinda Hodkiewicz[1], Michael Stewart[2], Tyler Bikaun[2]
1. Mechanical Engineering
2. Computer Science and Software Engineering
University of Western Australia

Simon Drew, Dave Belmont-Kleingeld
CEED Client: BHP Nickel West

Abstract

Maintenance Work Request Notifications (WRNs) are common data sources in industry that hold critical information for the reliability analysis of assets. A wealth of knowledge about common failure causes and effects remains untapped in the long-text description fields of these documents. Within even a small facility, there can exist many thousands of WRNs which makes manual processing a complex and arduous effort. This project uses Technical Language Processing (TLP) techniques to develop an automated and scalable solution for identifying causal information from unstructured WRN texts. A machine learning-based pipeline was developed to transform the unstructured text into a normalised format via a novel noise-removal method before extracting cause-and-effect relationships. Classification accuracy as measured by F1-score is 90% for the noise-removal task, and preliminary results show an accuracy of 76% for identifying causal relationships. Applying the pipeline to maintenance text improves the productivity of reliability engineers by efficiently providing insight into the causes of costly or chronic failures.

1. Introduction

Work Request Notifications (WRNs) are common data sources in industrial plants that hold useful information for reliability analysis and condition monitoring of assets. WRNs are generated each time maintenance work is required. They capture information about the state of an asset following inspections or malfunctions, and describe the corrective actions to be taken. Since 2010, over 150,000 WRNs have been recorded at the BHP NiW Kwinana (NWK) facility. The *long text* description field in each WRN maintenance record contains an average of 92 words, detailing information about failure modes and their subsequent effects. When logging fault descriptions, each technician describes failure events in their own way using free text. This results in rich, but challenging to align, texts making it difficult to consistently categorise failure modes. Further, the process for determining failure event histories is currently achieved by an operator manually analysing each maintenance record. Within even a small facility, there can exist many thousands of maintenance documents, which makes manual processing a complex and arduous effort. These challenges have made historical WRNs an underutilised data source, where valuable information such as failure causes and effects remains untapped.

Natural Language Processing (NLP) is a subfield of Artificial Intelligence for the automatic handling, comprehension and generation of human language by computers. NLP applications have changed people's daily lives with chatbots like Siri, and services like Google Translate, and have revolutionised business practices and operations. Recent work in the engineering domain has shown that care needs to be taken when applying out-of-the-box NLP techniques to technical domains. Technical unstructured data features jargon, misspellings and abbreviations. Further, many entities prevalent in domain-specific text do not exist in commonly used open knowledge bases, and hence input from domain experts is required. Brundage et al. (2021) proposed a framework for tailoring NLP solutions to industry use cases, namely Technical Language Processing (TLP). TLP is an expert-in-the-loop, iterative approach to adapting state-of-the-art NLP tools for technical languages and engineering use cases. This framework prioritises a continuous interaction between subject matter experts and technical analysts as they build TLP resources for maintenance-specific tasks. This research project aims to 1) convert *long text* maintenance documents into a normalised, machine-readable format via a novel noise removal method and 2) develop a TLP pipeline that identifies and extracts cause-and-effect relationships. Extracting causal information from maintenance records supports the construction of a graph database. This database assists technical personnel in efficiently and scalably querying the causes and effects of failure incidents across a multitude of WRNs, filtering on structured fields such as cost, date and asset type.

2. Process

The data flow process is illustrated in Figure 1. The pipeline first extracts the original maintenance records from the facility's Software, Applications and Products (SAP) Data Services platform. The data is then preprocessed and a Machine Learning (ML) model is applied to filter out problem-specific information. A second ML model is trained on annotated data and deployed on unseen WRNs to extract causal relationships and the items involved. The generated causal information is stored in a graph database, available for analysis from reliability engineers.

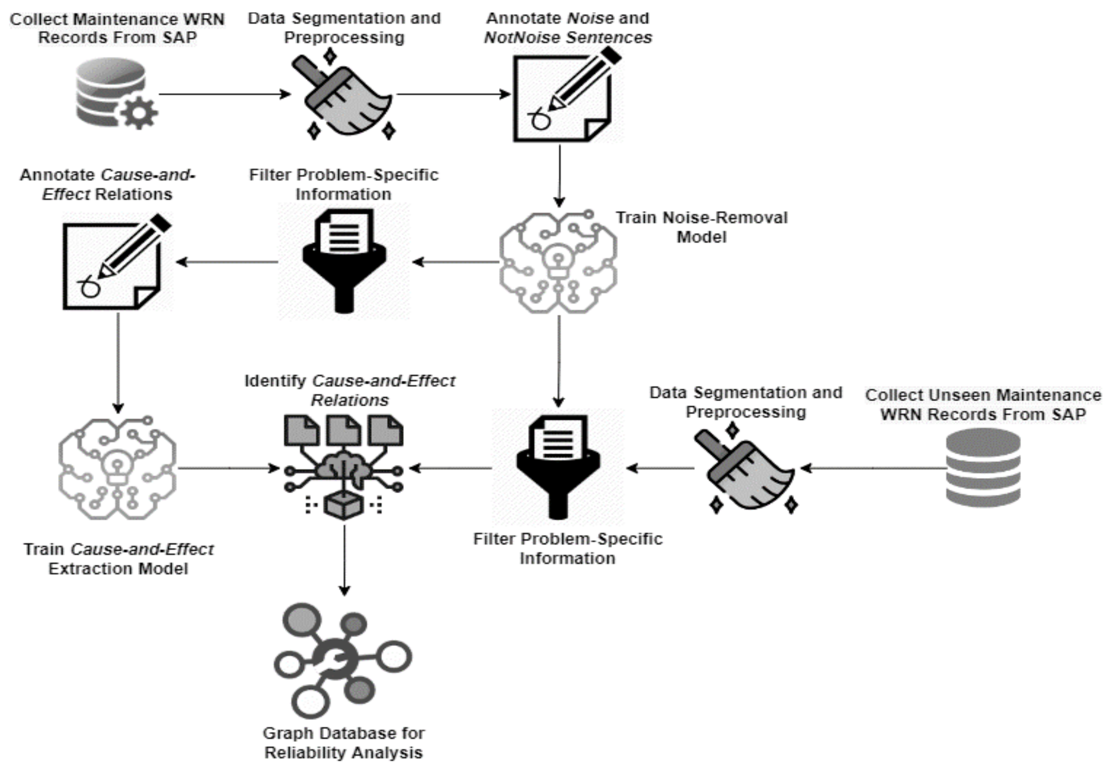


Figure 1 Process of the end-to-end pipeline for generating insights

2.1 Data Cleaning and Preprocessing

Preparing and preprocessing raw data into a standardised format is important for the performance of downstream NLP tasks such as cause-and-effect identification (Al Sharaou et al., 2021). A data segmentation and preprocessing module is written in Python to anonymise the maintenance data, correct spelling mistakes and normalise common domain-specific vocabulary to base forms.

2.2 Noise Removal using Problem-Relevant Classification

WRN *long-text* descriptions contain information superfluous to our cause-and-effect extraction problem such as equipment orders and time schedules. Compressing free-text documents into shorter extracts that retain problem-specific information is more effective input for downstream classification tasks (Pang & Lee, 2004). Applying this concept to the *long-text* maintenance data filters out problem-relevant information such as the failure mechanism, failure mode and items involved in a WRN while removing noisy data that is not relevant. Supervised ML requires annotated datasets, enabling computer algorithms to learn complex patterns and classify new data accurately. Annotations were performed on Redcoat, a web-based annotation tool (Stewart, et al., 2019). 1983 sentences from maintenance records were annotated as either *Noise* or *Not Noise* while consulting with subject matter experts. A state-of-the-art language model, Bidirectional Encoder Representations from Transformers (BERT), was trained on the annotated data to learn contextual, numerical representations of words (Devlin, et al., 2018). The embedding of a special token *[CLS]* is a representation of the entire sentence which is fed through to a classifier to determine if the sentence is *Noise or Not Noise*, as seen in Figure 2.

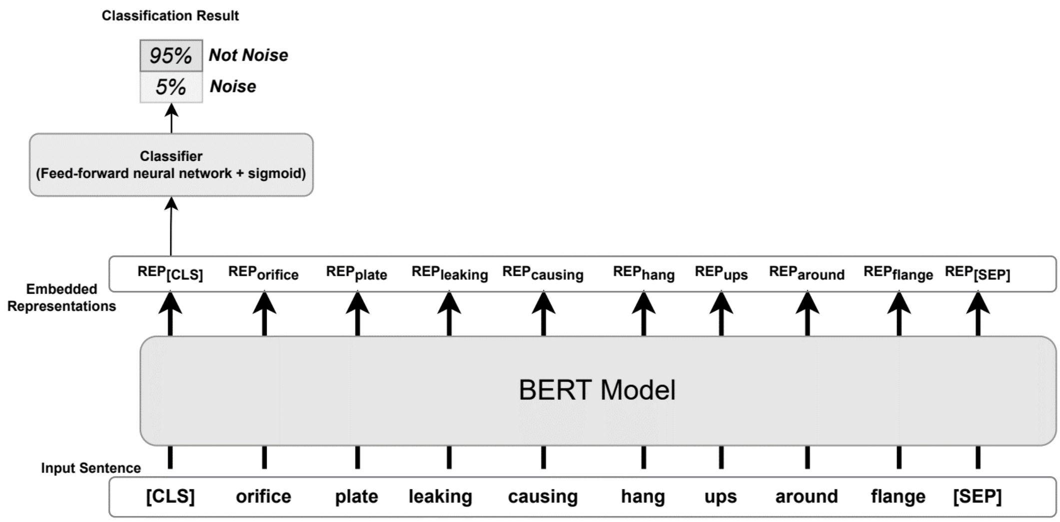


Figure 2 Sentence-level classification for problem-specific information

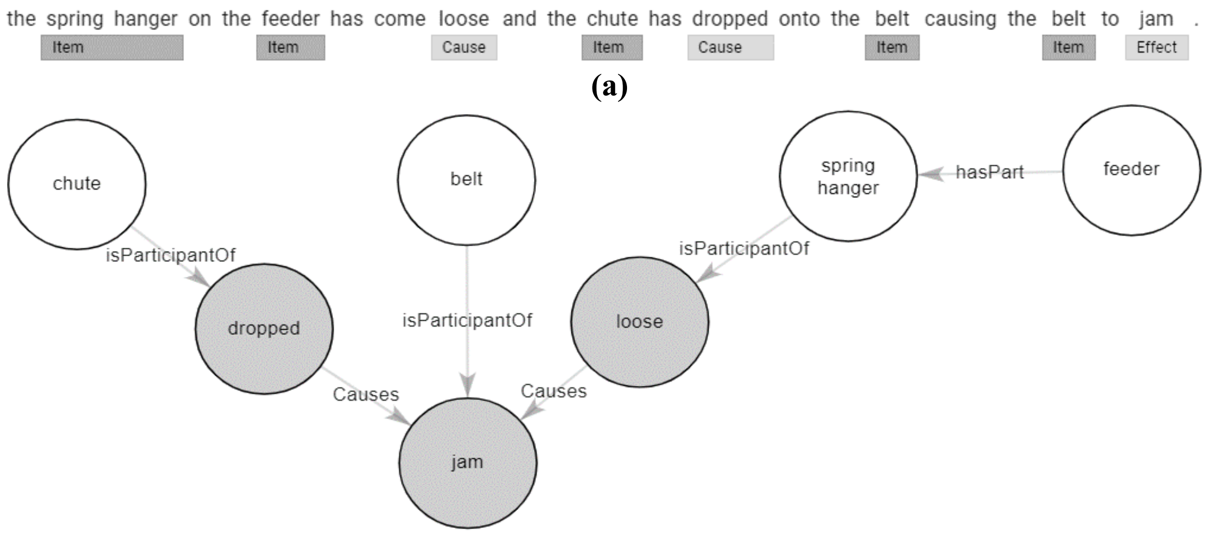
2.3 Machine Learning for Cause-Effect Identification

After the *long text* has been filtered using the noise removal model, a second ML model is trained to classify words as named entities and classify the relationships between them. 450 *long-text* documents are manually labelled entity and relation types by human annotators according to the maintenance-specific semantic concepts shown in Figure 3. These documents are split into training, validation and test sets with an 80:10:10 split, i.e. 80% of the data is used to train the model, 10% to validate the model and the final 10% as unseen data. The cause-and-effect identification model, again, uses BERT embeddings to learn contextual, numerical

Entity Type	Description	Examples
Item	A maintainable item or constituent part	pump, motor, bearing, handle
Cause	Physical, electrical, chemical, mechanical or other stresses that induce a problem, hazard or failure	wear, blown out, leak, dislodged
Effect	The consequences or impacts of causes on the system and/or component operation	jam, trip, overflow, high vibration levels
Relationship Type	**Description**	**Examples**
Causes	The occurrence of *entity 1* results in the occurrence of *entity 2*	'*blocking up* [Cause] Causes *high pressure* [Effect]'
isParticipantOf	The experience or state of an Item in an event	'*discharge line* [Item] isParticipantOf *split* [Cause]'
hasPart	A component, constituent, subpart or extension of an Item	'*clave* [Item] hasPart *vent* [Item] hasPart *air solenoid* [Item]'
Coreferences	Relationship between two *n-grams* in which both refer to the same thing	'*this* [Item] Coreferences *pump* [Item]'

Figure 3 Sentence-level classification for problem-relevance using BERT

representations of words. The model first classifies *Item, Cause* and *Effect* entities before linking entities together according to a particular relationship type. An example of the output for entity classification is shown in Figure 4 (a) and the corresponding relationship extraction is show in Figure 4 (b). The structure of these relationships is modelled based on an ISO 13379 standard causal tree which determines the root cause based on an existing set of failure modes (ISO 2003).



(a)

(b)

Figure 4 **(a)** Named Entity Recognition and **(b)** Relation Extraction model produces a causal tree

2.4 Graph Database for Decision Support Querying

The Information Extraction pipeline outputs structured relational data that can be stored in a graph database to facilitate interactive visualisation and querying of the WRNs. The graph produces colour-coded nodes according to entity type (*Item, Cause* and *Effect*) and displays edges representing the relevant relationship types. The graph database provides a platform for answering questions about common failure causes and effects at scale. Structured fields, including WRN creation date, asset functional location and maintenance cost, are added to the database to support queries seeking to trend data over time, location and cost.

3. Results and Discussion

The results demonstrate the efficacy of NLP techniques in transforming *long-text* maintenance documents into normalised data and extracting accurate cause-and-effect relationships for analysis. Standard precision, recall and F1-score are used as evaluation metrics to evaluate the accuracy of the trained models on unseen data. Precision calculates how many of the model's true predictions are actually true and recall calculates how many true examples the model predicts as true. F1-score is the harmonic mean of precision and recall. Table 1 outlines the noise removal classification results for WRN *long-text* data. The BERT classification model performs well in distinguishing between problem-specific sentences and noisy sentences.

Class	Precision	Recall	F1-Score
Noise	0.953	0.893	0.922
NotNoise	0.815	0.919	0.870
Weighted-average			**0.904**

Table 1 Results for sentence-level problem-relevance classification.

Table 2 displays the results for classifying causal relationships in the maintenance *long-text*.

Relationship Class	Precision	Recall	F1-Score
Causes	0.796	0.729	**0.761**
isParticipantOf	0.812	0.746	0.781
hasPart	1.000	0.667	0.800

Table 2 Results for extracting causal relationships between *Cause* and *Effect* entities. *isParticpantOf* and *hasPart* relation results are also shown.

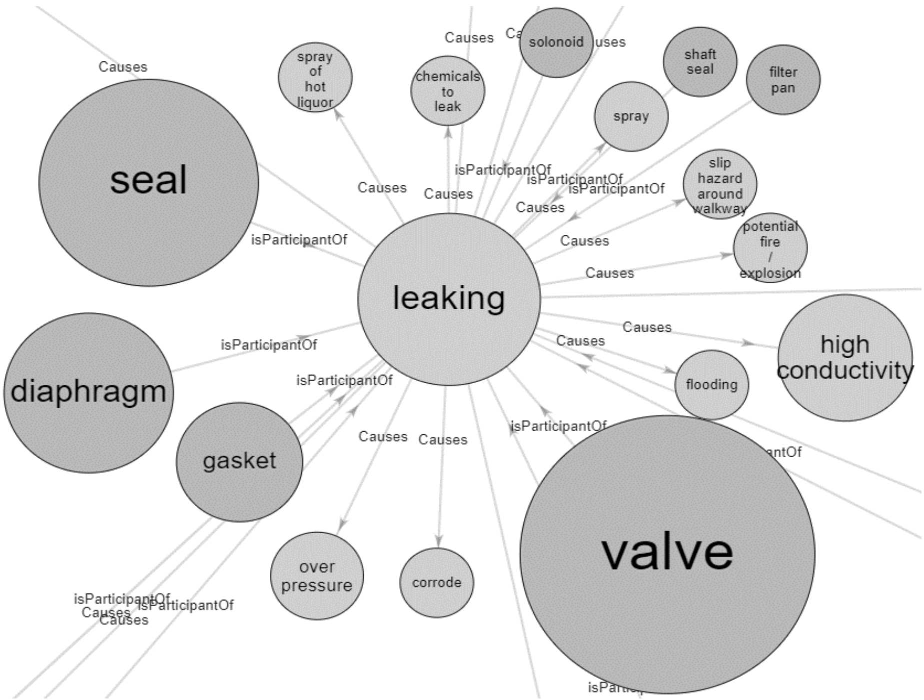


Figure 5 Example of frequently identified failure cause – leaking, and corresponding items and effects. Node size indicates relative frequency.

Visualising the graph database in the form of a knowledge graph yields insights into frequently occurring causes, their effects and the items involved. Figure 5 shows how filtering on large *Cause* nodes can produce insights. Leaking was identified as a commonly occurring cause as indicated by its relatively large node size. It is evident that leaking has caused various effects, including flooding, corrosion, overpressure and high conductivity. High conductivity is a commonly occurring effect linked with leaking, indicated by its relatively large node size. It is clear that valves, drains, diaphragms, and gaskets are all items frequently associated with leaks. Further examination reveals that leaks causing high conductivity are most strongly linked with diaphrams as opposed to valves or gaskets.

4. Conclusions and Future Work

The magnitude of information contained within *long-text* maintenance records demands an automated and scalable solution for processing this data. The end-to-end pipeline developed in this research demonstrates an effective method for transforming raw unstructured *long-text* into a structured output. The graph database provides an interactive querying platform to investigate questions such as: have Bredel hose pump failures increased over time at my facility? This information can produce insights for reliability engineers that can help guide their maintenance strategy. Classifying causal relationships was successful; however, classifying entities themselves is still an ongoing task. Training a computer model to recognise implicit causal information is an obvious challenge, and potential solutions such as incorporating domain knowledge are to be explored. In addition, future work could explore unlocking other knowledge embedded within WRN *long-text*, such as remedial information and safety insights.

5. Acknowledgements

This project would not have been possible without the support, mentoring and passion from my academic supervisors: Melinda Hodkiewicz, Tyler Bikaun and Michael Stewart, and my BHP NiW mentors Simon Drew and Dave Belmont-Kleingeld. I would also like to thank BHP NiW, Jeremy Leggoe and Kimberlie Hancock for providing this research opportunity.

6. References

Al Sharou, K., Li, Z., & Specia, L. (2021, September). Towards a better understanding of noise in natural language processing. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)* (pp. 53-62).

Brundage, M. P., Sexton, T., Hodkiewicz, M., Dima, A., & Lukens, S. (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, *27*, 42-46.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

ISO 13379:2003 (2003). Condition monitoring and diagnostics of machines — General guidelines on data interpretation and diagnostics techniques

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.

Stewart, M., Liu, W., & Cardell-Oliver, R. (2019, November). Redcoat: a collaborative annotation tool for hierarchical entity typing. In Proceedings of the 2019 Conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP): system demonstrations pp. 193-198.