

保全文書を対象とした故障分類情報の可視化手法の提案

二神 廉太郎[†]
株式会社 日立製作所[†]

室 啓朗[‡]
株式会社 日立製作所[‡]

1 はじめに

電力アセットなどの保守の現場では、日々の業務における保全の履歴を記録した保全文書が蓄積されている。保全文書は、事例ごとに観測された現象や実施した対策等の情報が網羅的に記載されており、その情報をデジタル技術により活用することで保守の効率化を行うことが期待できる。例として故障発生時に必要な対策を立案する業務では、現在と過去の事例の現象を抜け漏れなく比較する必要があり、保全文章を短時間で把握可能な可視化が求められている。

しかし保守の最適化を進める上で保全文書などのデータを収集し利用可能とする作業に全体の80%の工数が必要であると報告されている[1]。保全文書の情報抽出の障害として多数の専門用語や表記ゆれ、不統一なフォーマットなどがあり、汎用的な自然言語処理技術の適用が困難であった。保全文書を対象に情報抽出を行う従来研究では、語彙の意味情報を利用しないため誤抽出が生じる問題や[2]、抽出する故障イベントを人手で網羅的に定義するために大きな工数と専門知識が必要な問題があった[3]。

本研究では、保全文書の重要な情報である故障の現象と対策は、その情報と対応する語句を追うことで文書の大部分の内容を把握可能であることに着目した。こうした語句の定義を行うための固有表現抽出や類義語検出のタスクは、人手による少数の定義を学習し文書全体に定義する手法が既に確立されており[4]、低工数で利用できる。

本研究は、固有表現辞書を利用して保全文書から情報抽出し容易に把握可能な形態に可視化する手法を提案する。可視化形態として、故障の原因や依存関係を表すFT図[5]でなく、故障対策の立案に有用な故障の分類をツリー状に表現する、故障分類木と呼ぶ新形態を提案する。実験において本手法の情報抽出精度と文章の削減率を評価した。

2 提案手法

2.1 保全文章からの情報抽出

本研究で扱う保全文書は、故障事例ごとに発生した現象とその対策をフリーテキスト形式で記載した電子データである。保全文章から情報抽出を行うために、保全文書に出現する現象・対策とその対象部位を示す固有表現の辞書を利用する。その固有表現辞書として、語句を格納した word 列とその語句が持つ意味を定義した label 列を持った表形式の辞書を準備する必要がある。label 列に格納されるべき意味ラベルとして、部位を表す target、現象を表す observation、現象の否定を表す negative、対策を表す action のラベルを各語句に付与する必要がある。

最初の処理(1)として、保全文書の意味のまとまりごとに語句を検知し、故障分類木のグラフノード(以下、ノード)の単位で抽出する。事前に、保全文書の類義語を代表語で置換する処理を行う。抽出処理の具体的な手順として、文章の主語となる target ラベルの付与された語句(以下、target 語句)が1語以上現れたあと、その他のラベルの語句が出現した時にノードとする。このとき、ノードには target 語句と observation 語句または approach 語句のそれぞれを含む必要があり、approach 語句を含む場合は対策ノード、それ以外の場合は現象ノードと呼ぶ。各ノードは抽出元になった保全文書の情報を保持するものとする。

次の処理(2)として、ノードの語句の意味が類似するものを統合する処理を行う。そのために、2つの現象ノードの語句についてコサイン類似度を計算し、その類似度を集約したノード類似度が閾値以上の場合にノードを統合する。対策ノードも同様の統合を行う。negative 語句の有無が異なる場合にはノード類似度を0とする。具体的には、ラベル l の下で2つのノード A, B に対して式1に示す S_l を計算し、その結果を全ての l について平均した結果をノード類似度とする。ただし、 W_A はノード A に含まれる l 語句の集合で w_a はその元であり、 c はコサイン類似度である。

$$S_l(W_A, W_B) = \frac{\sum_{a \in A} \min_{b \in B} c(w_a, w_b) + \sum_{b \in B} \min_{a \in A} c(w_b, w_a)}{|W_A| + |W_B|} \quad (1)$$

最後の処理(3)として、故障分類木の可読性低下の要

A visualization of fault pattern extracted from maintenance report.

[†] Rentaro Futagami, Hitachi, Ltd. Research and Development Group.

[‡] Keiro Muro, Hitachi, Ltd. Research and Development Group.

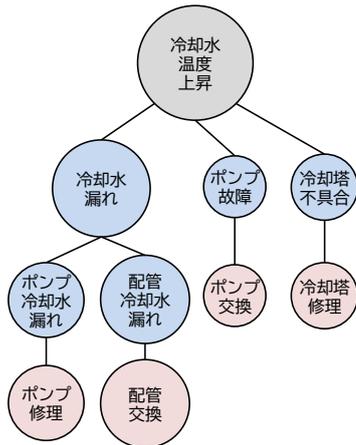


図1 冷却水温度上昇に関する故障分類木の例

因になる不必要な現象を削減する処理を行う。その削減方法として、故障現象からその対策を予測するランダムフォレストモデル [6] などを構築し、そのモデルにおける Gini 不純度 [6] 平均減少量などの変数の重要度を表す指標が小さい変数 (現象) を削除する。

2.2 故障分類木

2.1 節で抽出したノードをツリー状に配置し、故障分類木を構築する。故障分類木は図1に示すように、現象ノードを青色で、対策ノードを赤色で表示しており、木の根から葉までの一連のノードは現象と対策によって特徴づけられる保全文書の故障分類を表している。例として、図1の最も右端に位置する一連のノードは、「冷却水の温度が上昇し、冷却塔に不良が見つかった場合、その対策として冷却塔の修理を実施する。」という故障分類と対応する。故障分類木を用いることで、大量の保全文書の内容を横断的かつ迅速に把握可能である。

図1の根ノードとしては任意の現象・対策や保全文書に付与された故障コード等のメタ情報を指定し、現象ノードの場合はその現象ノードと同じノードが現れる保全文書を検索し、該当する保全文書の現象ノードをそのノードが出現する保全文書の数が多い順に根から分岐するように配置し、対策ノードも同様に配置することで故障分類木を構築できる。上記の処理においては、2.1 節の処理 (2) で統合されているノードは同じノードとして扱う。

3 評価実験

提案手法をガスエンジン保守に関する保全文書に適用し評価した。評価に際し、代表的な1件の故障コードについて正解の故障分類木 (以下、正解木) を専門家に作成頂き、提案手法によって自動的に構築した故障分類木と比較した。参考として、正解木1件の作成には20時間を要した一方で、故障コード全件 (104件) に対する故障分類木作成のためには、辞書の定義に10時間を要したのみ

表1 故障分類木と正解の故障分類木の比較結果

	保全文章	故障分類木	正解の故障分類木
対策再現率	90%	60% (67%)	100%
文章削減率	100%	14%	6%

であった。

具体的な評価指標として、正解木に記載された20種類の故障分類について、故障分類木において同じ分類を再現できた割合を対策再現率として計算した。また、内容把握の容易さを評価するために、文章の量 (ノードの数) を削減した割合を文章削減率として計算した。その結果を表1に示す。ただし故障コードの誤りなどのため、正解木に記された現象と対策の全てが保全文章に記載されていないため、保全文書比の対策再現率を括弧付きで表記した。

表1を見ると、故障分類木において文章削減率は良い結果が得られ、保守の実務者からも定性的に高い評価を得た。一方で、対策再現率の指標においては改善の余地が見られた。本実験における以上の結果から、提案手法は保全文書の内容把握時間の短縮する効果があり、少ない工数で利用可能であることが認められた。

4 おわりに

本研究は保全文書の現象と対策を短時間で把握可能とする情報抽出と可視化の手法を提案した。本手法を用いて低工数で構築可能な故障分類木と呼ぶ可視化を利用することで、故障の対策立案を短時間で行うことが可能である。

参考文献

- [1] 新井 信行：予兆検知システムとデータおよび分析駆動型「戦略的メンテナンス」、火力原子力発電, Vol.7, No.2, pp.186-190 (2019).
- [2] Farley, B.: Extracting information from free-text aircraft repair notes, *AI EDAM*, Vol.15, pp.295-305 (2001).
- [3] 安藤 英幸ほか：テキストマイニングを用いた故障報告書分析手法の研究, 日本造船学会論文集, Vol.192, pp.475-483 (2002).
- [4] Zhang, C.: DeepDive: A Data Management System for Automatic Knowledge Base Construction, *Univ. of Wisconsin-Madison Thesis*, (2015).
- [5] Lee, W. S., et al.: Fault Tree Analysis, Methods, and Applications – A Review, *IEEE trans. on reliability*, Vol.34, No.3 pp.194-203 (1985).
- [6] Breiman, L.: Random forests, *Machine Learning*, Vol.45, No.1, pp.5-32 (2001).