

保守文書に記載される故障表現間の関係性抽出技術の開発

Relation Extraction between Failure Expressions in Maintenance Documentation

我妻 正太郎[†]
Shotaro Agatsuma照屋 絵理[†]
Eri Teruya竹内 理[†]
Tadashi Takeuchi

1. はじめに

医用機器メーカー等の製造業では国際的な製品品質向上・画一化に伴い、顧客価値に繋がるサービスをデジタル技術を用いて提供することにより、製品の付加価値向上をめざす動きが盛んである。上記サービスの一例として、製品の保守手順推薦を行うリペアレコメンドサービスが挙げられる。このサービスは製品に起こった現象を入力すると、その現象に対する原因や適切な処置を提示するものである。サービスの実現には、製品に起こった現象とその原因、処置の関係性を示したデータ（以下保守ナレッジと呼称）を準備しておくことが必須である。保守ナレッジは製品保守現場における日々の保守作業を記録した保守報告書から抽出できる。しかし、報告書は非定型かつ専門的な表現を多く含むため、人手による保守報告書からの保守ナレッジ抽出には深い専門知識と膨大な時間が必要である。よって、保守ナレッジ自動抽出技術の開発がリペアレコメンドサービス実現には重要である。

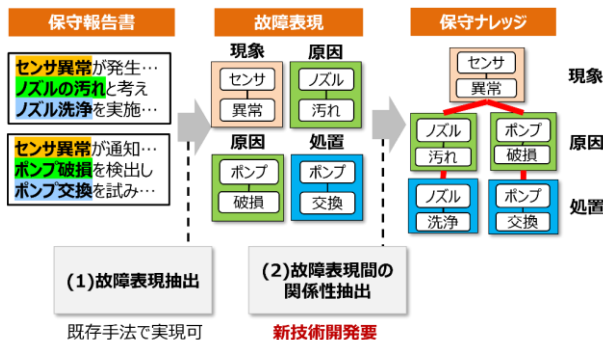


図1 保守ナレッジ抽出概要図

そこで、本研究では保守ナレッジ自動抽出技術を開発するため、関係付けるべき現象・原因・処置を対象とする部品名の部品構成表 (Bill Of Material, BOM) 上の距離が近いもしくは一致する、という保守作業の特徴を用いた関係性抽出技術を提案する。さらに、提案手法を用いて保守ナレッジ抽出を行い、その精度を既存手法と比較した。その結果、提案手法は文書が非定型かつ少量である場合における保守ナレッジ抽出精度向上に有用であることが示唆された。

2. 保守ナレッジ抽出の課題

抽出すべき保守ナレッジは、図1右図のような現象・原因・処置の関係性を示したグラフ構造データで表現できる。グラフデータを構成するノードは、「異常」のような現象・原因・処置を表す単語と、「センサ」のような現象・原因・処置が対象とする部品名のペアで構成される。以下

[†](株)日立製作所 Hitachi, Ltd.

ではこのペアを「故障表現」と呼称する。保守報告書の記載から、故障表現同士を文意に合うように関係づけたグラフ構造データが保守ナレッジとなる。想定するリペアレコメンドサービスでは、まず製品に起こった現象の故障表現をユーザから受け取る。その後現象に対応する保守ナレッジ内の原因の故障表現と、さらに原因の故障表現に対応する処置の故障表現を検索し、ユーザに検索結果を提示する。

保守ナレッジ自動抽出には図1に示す通り、(1)報告書からの故障表現抽出と、(2)故障表現間の関係性抽出の2ステップが必要である。(1)は既存の固有表現抽出技術 [1]を用いて抽出した固有表現同士を、文章中の係り受け関係を用いて関連付けることで実現できる。一方、(2)を既存の関係性抽出手法で実現するのは難しい。

関係性抽出の既存手法では、例えば保守報告書中の「<原因>のため<処置>」のような原因と対処を関係づける特徴的な周辺語パターンを参考に表現間の関係付けを行う。具体的には、機械学習により文書から周辺語パターンを学習したモデルを作成する方法 [2]や、周辺語パターンを参考に作成したルールにより関係性抽出を行う手法 [3]が存在する。機械学習手法の適用には、特徴的な周辺語パターンが記載された大量の学習用文書が必要となる。しかし、保守報告書等の社内文書では周辺語パターンを十分に学習できるほどの文書量が無い場合が多い。さらに報告書は作成者によって文章のフォーマットが異なる場合があり、特徴的な周辺語パターンのルール化も難しい。

これらの問題により、既存手法を保守報告書に適用しても故障表現関係付け精度は低いと考えられる。例えば、既存手法 [3]を医用機器関係の保守報告書に適用した場合の故障表現関係付け精度は約36%であり、精度は低い。よって、報告書が非定型かつ数が少ない場合における(2)の精度向上が課題となる。

3. 課題解決手法

本研究では、医用機器関係の保守報告書を対象に課題解決手法を検討した。課題解決にあたり、各故障表現には部品名という同じ性質の単語が含まれる点に着目し、部品名の関係性から故障表現同士の関係付けを行えないか検討した。保守報告書を調査した結果、関係付けるべき故障表現の間には2点の特徴があることが分かった。一つは「ノズル汚れ」「ノズル洗浄」のように、同じ部品名を持つ故障表現同士が関係付くことが多い点である。保守作業では「ノズル汚れ」に対して「ノズル洗浄」を行うように、現象・原因・処置が対象とする部品名が共通している場合が多いため、この特徴に着目した。もう一つは装置上の物理的距離が近い部品名同士を持つ表現同士が関係付くことが多い点である。製品の構成上、故障が起こった部品の近くにある別の部品が故障原因になっている場合が多く、この特徴にも着目した。これら2点の特徴を考慮したルールによって故障表現関係付けを行う。

上述の 2 種類の特徴を用いた故障表現関係付け手法を、図 2 に示す具体的な例を用いて説明する。例えば、あらかじめ既存手法により故障表現が抽出できていると仮定する。まず、類義語となる部品名を持つ故障表現同士の関係付けを行う。ここでは「センサ」「センサー」が類義語であるため、これらの単語を持つ故障表現同士を関係づける。部品名同士の類義語判定は、機械学習による既存手法 [4] を用いる。次に製品の BOM 情報 (図 2 右下「BOM」) を参照し、故障表現同士に含まれる部品名が同じモジュールに含まれるものを関係づける。ここでは「センサ」「ノズル」が同じモジュールに含まれるため、これらの単語を持つ故障表現同士を関係づける。このような 2 つの処理によって、少ない文書量かつ非定型の文書からでも高精度に故障表現関係付けを行う。

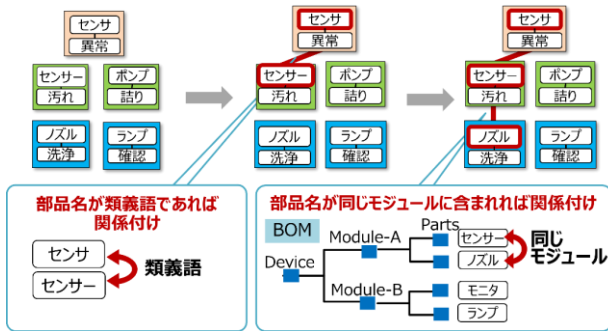


図 2 提案手法概要図

4. 評価

提案手法と、日本語非定型文書からの関係性抽出に関する既存手法を比較し、既存手法における問題となっていた必要文書量と関係性抽出精度がどれだけ改善したか評価した。本評価では、提案手法、ルールベースによる既存手法 [3]、機械学習による既存手法 [2] の 3 点を比較した。ルールベースによる既存手法では、文章中の表現の出現順を用いたルールにより、保守報告書中の表現間の関係性抽出を行う。機械学習による既存手法では、機械学習モデル BERT [5] に大量の文書を入力し、各表現間を関係付ける特徴を学習させることで関係性抽出を行う。

比較時の条件を以下に示す。提案手法により医用機器関係の報告書 50 件中の現象・原因・処置間の関係性抽出を実施した。抽出した結果から関係性抽出の正答率を算出し、この正答率を関係性抽出精度とした。提案手法及びルールベース手法では、故障表現抽出のため、4,000 件の医用機器関係の報告書を用いた。一方、既存機械学習手法については、学習の前提となる現象・原因・処置の関係がラベル

表 1 既存手法と提案手法の性能比較結果

評価尺度	関係性抽出手法		
	既存ルールベース手法 [3]	既存機械学習手法 [2]	提案手法
精度 (正解率)	36.4%	71.4%	65.2%
必要文書量	4,000 件 (保守報告書)	約 1,000,000 件 (wikipedia 全文)	4,000 件 (保守報告書)

付けされた報告書を用意するには膨大な時間がかかるため、今回は仮に論文 [2] 記載の関係性抽出精度を比較対象として用いた。

比較結果を表 1 に示す。提案手法により、ルールベース既存手法の精度を約 30% 上回る抽出精度を、既存機械学習手法の約 1/2,500 の文書量で達成できた。上記の結果から、提案手法は文書が非定型かつ少量である医用機器関係の保守報告書からの保守ナレッジ抽出精度向上に有用であることが示唆された。一方で、提案手法は既存機械学習手法より精度が低く、精度改善が課題となる。提案手法による関係性抽出失敗時の原因は、現象・原因・処置が対象とする部品名の記載が省略されており故障表現が未抽出になる場合と、提案手法で設定したルールが故障表現同士の部品名に対応していない場合の 2 点である。前者の解決には、省略されている部品名の情報を外部から入力することで故障表現の未抽出を減らせる可能性がある。例えば医用機器関係の保守報告書中で省略された部品名は、報告書記載の作業に使用した部品をまとめた作業来歴書に記載されている場合があり、作業来歴書記載の部品名と保守報告書中の現象・原因・処置名を照合することで未抽出を減らせる可能性がある。後者は提案手法での対処が難しく、このような場合における関係性抽出精度向上には、提案手法と他の手法を複合して用いる必要がある。

5. おわりに

本研究では、製造業向けリペアレコメンドサービスの実現のため、保守報告書からの保守ナレッジ自動抽出技術を開発した。なお、保守ナレッジは製品に起こった現象とその原因、処置の関係性を示したデータとする。

本研究では関係付けるべき現象・原因・処置は対象とする部品名の BOM 上の距離が近いもしくは一致する、という保守作業の特徴を用いたルールによる関係性抽出技術を開発した。上記技術により、報告書が非定型かつ少量である場合にも高精度に保守ナレッジ抽出を実現できる。開発手法と既存関係性抽出手法を比較したところ、精度は約 30% 上昇し、必要文書量は約 1/2,500 に削減された。以上から、本手法は、文章が非定型かつ少量である保守報告書からの保守ナレッジ抽出精度向上に有用であることが示唆された。今後は更なる精度向上を図る。さらに、医用機器関係とは別種類の製品に関する文書に本手法を適用し、高精度な関係性抽出可能か調査する。

参考文献

- [1] 照屋絵理, 愛甲和秀, 竹内理, “Smart Dictionary 実用化に向けた教師データ量と NER 精度評価について”, 第 19 回情報科学技術フォーラム (2019).
- [2] 高津弘明, 安藤涼太, 松山洋一, 小林哲則, “会話によるニュース記事伝達のための談話構造解析”, 言語処理学会第 27 回年次大会発表論文集, pp. 966–970, (2021).
- [3] 二神 廉太郎, 室 啓明, “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”, 第 83 回情報処理学会全国大会講演論文集, pp. 43–44, (2021).
- [4] 我妻 正太郎, 照屋 絵理, 竹内 理 “自動作成された類義語抽出ルールによる類義語抽出精度向上手法”, 第 21 回情報科学技術フォーラム (2021).
- [5] Jacob Devlin et al. “BERT: Pre-training of deep bidirectional transformers for language understanding”. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186, (2019)