

設備保全レポート活用に向けたテキスト構造化の検討

高橋 拓誠 谷口 元樹 谷口 友紀 大熊 智子
 富士フイルム株式会社
 takumi.c.takahashi@fujifilm.com

概要

本稿では、設備保全レポートを対象としたテキスト構造化の検証を行う。具体的には、文分類、固有表現抽出、モダリティ判定、関係抽出の4つのタスクを対象に、事前学習済みモデル BERT を用いたテキスト構造化の評価を実施した。個別のタスクの評価では、どの構造化タスクでも高精度に構造化可能であることを確認した。さらに、より実際の活用に応じた検証として、同一の保全レポートに対して複数の構造化タスクを適用した場合の性能評価と分析を行い、実運用における課題について議論する。

1 はじめに

製造現場における日々の生産プロセスでは、設備にトラブルが発生した際に迅速な対応が求められる。こうした現場の保全業務は、過去の異常や処置などについて記録された保全レポートを参考にしながら対策を検討することも多く、膨大に蓄積された保全レポートの活用が現場で求められている。保全レポートは、一般的に自然文で記述されるため、記載された内容をあらかじめ構造化しておくことで効果的に情報検索やデータ分析に活用することができる。保全レポートを対象としたテキストの構造化では、テキスト分類 [1, 2], 固有表現・フレーズ抽出 [3, 4, 5], 原因推定 [6] などの研究がなされてきた。

一方で、実際の保全レポートの活用を考えると、複数の構造化情報を同時に扱いたい状況が起り得る。例えば、“異常が発生した部品”を対象に分析を行いたい場合、あらかじめ異常に関して記載された文を構造化しておき、それらの文を対象に部品名を抽出するのが望ましい。したがって、保全レポートを活用するために、同一文書に対して複数の構造化タスクを検証することは重要であるといえる。

富士フイルムエンジニアリングでは、保全レポート、設備情報、日々の稼働記録などを一元管理する

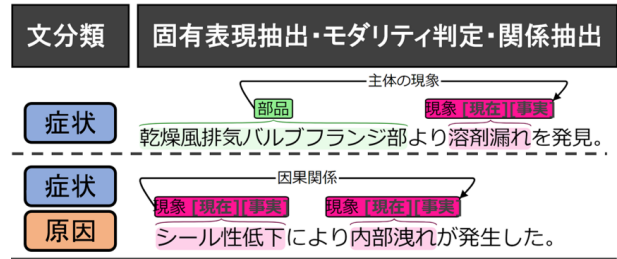


図1 構造化された保全レポートの例。4種類の構造化ラベルが付与される。

ためのシステム KARTEMIX を開発している。我々は、本システムに登録された保全レポートを用いて、文分類、固有表現抽出、モダリティ判定、関係抽出に関する複数のタスクについてアノテーションしたデータセットを構築した [7]。図1に構造化された保全レポートの例を示す。

本稿では、保全レポートを対象に複数の構造化タスクに対してベースラインを検証した結果について報告する。具体的には、文分類、固有表現抽出、モダリティ判定、関係抽出の各構造化タスクについてアノテーションされた保全レポートを対象に、各構造化タスクに対して個別にベースラインを評価した。さらに、より実践的な例として、保全レポート全体に対して同時に複数の構造化タスクを適用した場合の性能評価とその分析結果について議論する。

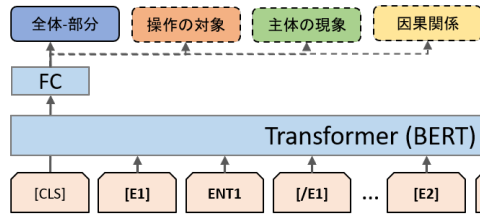
2 テキスト構造化

本研究では、事前学習済み言語モデルである BERT [8] を各構造化タスクに適用する。図2に各タスクのモデルの概要を示す。以降、文分類 (2.1 節)、固有表現抽出・モダリティ判定 (2.2 節)、関係抽出 (2.3 節) のモデルについて個別に説明する。なお、各ラベルの詳細は、谷口ら [7] の研究を参照されたい。

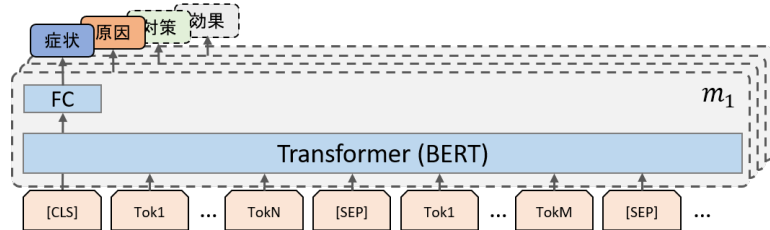
2.1 文分類

文分類では、 N 個の文 $\{x_1, \dots, x_N\}$ から構成される文書を対象に、予測対象の文 x_i が **症状**, **原因**, **対**

(c) 関係抽出



(a) 文分類



(b) 固有表現抽出・モダリティ判定

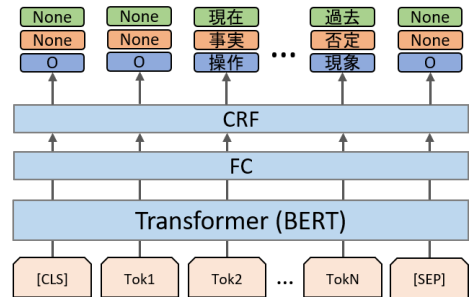


図2 各タスクにおける構造化モデルの概要.

策, 効果のいずれに該当するかマルチラベル分類する. 具体的には, 予測対象のクラス $c \in \{ \text{症状}, \text{原因}, \text{対策}, \text{効果} \}$ の有無を二値判定するモデル m_c をあらかじめ学習しておき, 推論時に各モデルを x_i に対して個別に適用する.

学習および推論では, 予測対象 x_i および x_i 以降に記載された文章 $\{x_{i+1}, \dots, x_{i+j}\}$ を同時にモデルへ入力する. 形式的には, $[\text{CLS}]x_i[\text{SEP}]x_{i+1}[\text{SEP}], \dots, x_{i+j}[\text{SEP}]$ をモデル m_c へ入力して, d 次元の分散表現 \mathbf{h}_i を得る. なお, \mathbf{h}_i は $[\text{CLS}]$ の分散表現に対応する. 最終的に, 分散表現 \mathbf{h}_i を全結合層 (FC) へ入力し, x_i に対するクラス c の有無を表わすラベル y_i^c を判定する.

2.2 固有表現抽出・モダリティ判定

固有表現抽出では, 文に含まれる各単語に対して, 固有表現ラベル $e \in \{ \text{部品}, \text{操作}, \text{現象}, \text{値}, \text{日時} \}$ を付与する. モダリティ判定では, 現象と操作をあらわす固有表現に対して, 事実性ラベル $f \in \{ \text{事実}, \text{否定}, \text{その他} \}$ と時制ラベル $t \in \{ \text{現在}, \text{過去}, \text{未来} \}$ をそれぞれ判定する.

本研究では, 固有表現抽出およびモダリティ判定のラベルを統合し, 1つのモデルで同時に抽出を行う. 具体的には, L 個の単語から構成される対象文 $x_i = \{w_1, \dots, w_L\}$ をモデルへ入力し, 単語 w_k について分散表現 \mathbf{h}_k を得る. 続いて, 分散表現 \mathbf{h}_k を全結合層 (FC) へ入力し \mathbf{h}'_k を得る. 最終的に, \mathbf{h}'_k を CRF 層 [9] へ入力して単語 w_k に対する固有表現ラベル y_k^e , 事実性ラベル y_k^f , 時制ラベル y_k^t を判定する.

2.3 関係抽出

関係抽出では, 文 x_i に含まれる固有表現 1 (*sub*) と固有表現 2 (*obj*) の間に付与される関係 $r \in \{ \text{全体部分}, \text{操作の対象}, \text{主体の現象}, \text{因果関係} \}$ を推定する. 本稿では, Entity Marker に基づく方法 [10] に従い, 以下の2種類の入力方式を採用した.

Simple-marker は, *sub* と *obj* の前後に, 対象の固有表現の位置を認識するための特殊トークンを挿入する. 具体的には, $\dots, [\text{E1}]sub[/\text{E1}], \dots, [\text{E2}]obj[/\text{E2}], \dots$ と表記する.

Entity-type は, *sub* と *obj* の前後に, 対象の固有表現の位置および固有表現ラベルを認識するための特殊トークンを挿入する. 具体的には, 固有表現ラベル $e1$ をもつ *sub* と固有表現ラベル $e2$ をもつ *obj* に対しては, $\dots, [\text{E1}_e1]sub[/\text{E1}_e1], \dots, [\text{E2}_e2]obj[/\text{E2}_e2], \dots$ と表記しモデルへ入力する.

上記の入力に対して得られた d 次元の分散表現 \mathbf{h}_i を全結合層 (FC) へ入力して関係ラベル y_i^r を推定する. 文分類と同様に, \mathbf{h}_i は $[\text{CLS}]$ の分散表現に対応する. なお, *sub* と *obj* の間にいずれの関係も付与されない場合は, “関係なし” ラベルを推定する.

3 評価実験

本節では, 各構造化タスクにおけるベースラインの性能を個別に検証する. まず, 個別のタスクの評価を行うことで, 保全レポートに対するテキスト構造化の実現性や各タスクの難易度を明らかにする.

表1 文分類の実験結果 (F 値).

モデル	症状	原因	対策	効果	平均
BERT _{base}	0.753	0.827	0.821	0.680	0.770
+ context	0.761	0.834	0.838	0.759	0.798

実験では、保全レポートを対象に複数のタスクのアノテーションを付与したデータセット [7] を 8:1:1 の割合で訓練・開発・評価に分割して使用した。

3.1 文分類タスク

文分類タスクでは、2.1 節のモデルに基づき、2 種類のベースラインを用いた検証を行う。具体的には、分類対象の文のみ入力するモデル (BERT_{base}) と、分類対象の文とそのあとに記載された文章も同時に入力するモデル (BERT_{base}+context) を比較する。モデルの訓練では、学習率を $2e-5$ に設定し、Adam による最適化を行った。バッチサイズは 32、エポック数は 3、入力の最大長は 512 に設定した。

表 1 に文分類の実験結果を示す。対象文のみ入力した BERT_{base} と比較して、対象文より後ろに記載された文章も同時に入力した BERT_{base}+context の方が性能が高いことが分かる。特に、効果クラスにおいて性能の差が顕著であることが分かった。これは、対策の効果を判断するための手がかりが、分類対象の文に後続することが多いためである。

3.2 固有表現抽出・モダリティ判定タスク

固有表現抽出およびモダリティ判定タスクでは、2.2 節のモデルに基づき、2 種類のベースラインを用いた検証を行う。具体的には、BERT に全結合層を追加したモデル [8] と、BERT に全結合層と CRF 層を追加したモデル (図 2) を比較する。どちらのモデルにおいても BERT_{base} を利用し、入力の最大長は 128、バッチサイズは 32、エポック数は 3、学習率は $3e-5$ に設定し、Adam による最適化を行った。

表 2 に固有表現抽出の実験結果を示す。いずれの固有表現カテゴリにおいても、BERT と比較して BERT+CRF の方がより高い性能を示している。

表 3 に事実性判定の実験結果を示す。事実性判定では、“事実”と“否定”カテゴリにおいて、固有表現抽出と同等の性能を示した。一方で、“その他”カテゴリでは、性能が著しく低下した。これは、仮定/条件のように、事実性を一意に定められない表現が多く含まれることや (4.2 節)、“その他”カテゴリが訓練データで頻出しないためだと考えられる。

表2 固有表現抽出の実験結果 (F 値).

モデル	部品	操作	現象	日時	値	全体
BERT	86.0	84.8	83.4	86.7	75.7	83.9
BERT+CRF	88.6	87.1	86.0	89.8	81.2	86.8

表3 事実性判定の実験結果 (F 値).

モデル	事実	否定	その他	全体
BERT	85.7	79.9	26.7	82.7
BERT+CRF	87.6	77.8	36.4	84.3

表 4 に時制判定の実験結果を示す。“現在”カテゴリでは、高い性能を示したものの、“未来”と“過去”カテゴリで性能が著しく低下することが分かった。理由として、上記 2 クラスは訓練データでほとんど出現しないクラス¹⁾であることが挙げられる。

3.3 関係抽出タスク

関係抽出タスクでは、2.3 節に記載した 2 種類のベースライン (Simple-marker, Entity-type) を比較する。いずれも BERT_{base} を事前学習済みモデルとして利用した。訓練では、入力の最大長を 128、バッチサイズを 32、エポック数を 3、学習率を $2e-5$ に設定し、Adam による最適化を行った。なお、対象の固有表現の前後に挿入する特殊トークンの分散表現も同時に fine-tuning した。さらに、対象の固有表現ラベルの組み合わせが関係ラベルの条件を満たす事例をすべて正例として扱う Heuristics を用意した。

表 5 に関係抽出の実験結果を示す。Heuristics より、“操作の対象”や“主体の現象”は正例の割合が比較的高く識別が易しいクラスである一方、“全体-部分”や“因果関係”は正例の割合が低く、識別が難しいクラスであることが分かる。Simple-marker および Entity-type は、いずれのクラスにおいても Heuristics の結果を大きく上回る性能を示した。また、BERT を用いたベースラインの識別精度が Heuristics の傾向と一致することから、データセットの正例と負例のバランスが識別の難しさを表わしていることが分かる。因果関係のような識別精度が低いクラスについては、データの不均衡性を緩和させる学習方法を採用するなど、いくつかの改善が必要である。

4 複数タスクによる構造化の分析

3 節では、各構造化タスクのベースラインを個別に評価した。一方で、個別のタスクに対する評価だ

1) 訓練データにおける時制ラベルの割合は、現在 = 90%、過去 = 4.6%、未来 = 5.4% であった。

表4 時制判定の実験結果 (F 値).

モデル	現在	過去	未来	全体
BERT	85.7	26.0	61.5	83.2
BERT+CRF	87.6	24.5	64.4	85.1

表5 関係抽出の実験結果 (F 値). H: 全体-部分, O: 操作の対象, S: 主体の現象, C: 因果関係を示す.

モデル	H	O	S	C	平均
Heuristics	0.335	0.770	0.635	0.084	0.456
Simple-marker	0.780	0.968	0.841	0.626	0.804
Entity-type	0.777	0.970	0.863	0.617	0.807

けでなく、レポート全体の構造化がどの程度可能なかの議論することは、検索や分析に活用するうえで重要である。本節では、保全レポートに対して複数の構造化タスクを適用した結果について分析する。

4.1 レポート全体の構造化の評価

レポート全体に対する複数の構造化タスクの性能を評価する。文分類は BERT_{base+context}(表1), 固有表現抽出とモダリティ判定は BERT+CRF(表2-4), 関係抽出²⁾は Simple-marker(表5)を用いて検証した。

図3に各構造化タスクにおいて閾値 (F 値) を超えた保全レポートの件数の推移を示す。各構造化タスクを比較すると、固有表現抽出・モダリティ判定や関係抽出は、レポート全体の構造化の精度が比較的高いことが分かる。一方で、文分類は他のタスクと比較して構造化の精度が低いレポートが多いことが分かった。これは、表1より、**症状**や**効果**の構造化の精度が十分でないことや、他の構造化タスクと比較して、レポート中の予測対象の数が少なく推定誤りによる影響が大きいためである。また、すべての構造化タスクにおいて、F 値 0.7 を超えたレポートは全体の約 24%であった。一方で、全ての構造化タスクで F 値 0.7 を下回るレポートは全体の約 7%であった。その他のレポート (69%) においては、各タスクの性能の傾向がレポートごとに異なっていたことから、タスク間の知識を共有しながら学習する仕組みを採用するなど、いくつか改善が必要である。

4.2 構造化のエラー事例

文分類: 対策/効果の文に誤分類される事例が散見された。具体的には、“検討”や“確認”のように、症状の解決に直接寄与しない対応が対策/効果を表わす文として誤分類される傾向にあった。

固有表現抽出: 固有表現抽出では、**現象と操作の**

2) 予測対象の2つの固有表現は、正解データを使用した。

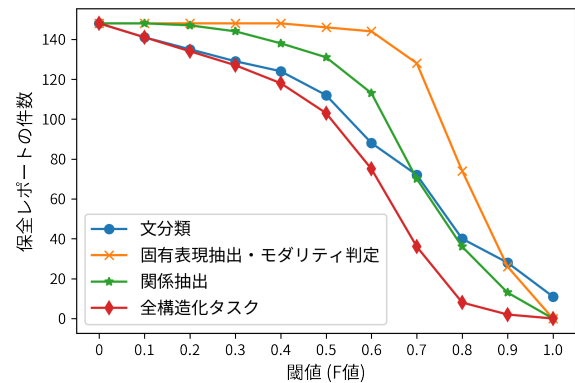


図3 閾値 (F 値) を超えた保全レポートの件数の推移。各タスクの F 値はレポート中の平均 F 値、全構造化タスクの F 値は各構造化タスクの平均 F 値の最小値とした。

ラベル間で誤分類が発生していた。例えば、“復旧”や“切替”という単語は、人が設備に対して実行した動作なのか、部品の状態を表わす単語なのかによって固有表現ラベルも異なるため、各単語が文脈上で持つ機能に関する判断が必要である。

モダリティ判定: 事実性判定では、仮定/条件のように、事実性を一意に定められない記載に対する誤分類が散見された。具体的には、“～で故障すると流延停止に至る”のように、一般的な知識や特定の状況を仮定した場合の記載が該当する。このような場合は、**その他**に分類されるべきだが、BERT+CRFの推論では**事実**に分類されていた。

関係抽出: 全体-部分において逆向きの関係を誤検出する事例が散見された。具体的には、“コントローラーボックスの基盤”という記載では、コントローラーボックスの部品として基盤が抽出されるべきであり、その逆は成立しない。

5 おわりに

本研究では、設備保全レポートを対象とした複数のテキスト構造化タスクに関する性能評価を実施した。具体的には、文分類、固有表現抽出、モダリティ判定、関係抽出の4つのタスクについて評価し、個別の評価では高精度に構造化可能なことを確認した。一方、複数の構造化タスクを保全レポートに同時に適用した評価では、すべてのタスクで高精度に構造化可能なレポートは多くないことが分かった。

今後の課題として、保全レポート全体の複数の構造化情報をより高い精度で抽出可能にすることが挙げられる。さらに、本稿で検証した構造化情報を検索などに適用し、その効果を検証する予定である。

商標

KARTEMIX は富士フイルムエンジニアリング株式会社の登録商標です。

参考文献

- [1] 勝又智, 小町守, 真鍋章, 谷本恒野. 障害レポートの分類問題に対するデータ選択を用いた BERT モデルの精度向上. 言語処理学会 第 26 回年次大会, pp. 645–648, 2020.
- [2] 山下郁海, 小町守, 真鍋章, 谷本恒野. 隠れ層補間によるデータ拡張を用いた障害レポート分類. 言語処理学会 第 27 回年次大会, pp. 1794–1798, 2021.
- [3] 小平知範, 宮崎亮輔, 小町守. 障害情報レポートに対する同時関連文章圧縮. 言語処理学会 第 23 回年次大会, pp. 186–189, 2017.
- [4] 伊藤雅弘, 山崎智弘. アノテーション漏れ推定を用いたエンティティ抽出. 言語処理学会 第 27 回年次大会, pp. 1264–1268, 2021.
- [5] 本間広樹, 小町守, 真鍋章, 谷本恒野. BERT モデルを用いた障害レポートに対する重要箇所抽出. 言語処理学会 第 27 回年次大会, pp. 189–193, 2021.
- [6] 勝又智, 小町守, 真鍋章, 大頭威, 嶋崎優子. node2vec を用いた障害レポートにおける故障原因推定. 言語処理学会 第 25 回年次大会, pp. 1045–1048, 2019.
- [7] 谷口元樹, 高橋拓誠, 谷口友紀, 大熊智子. 設備保全レポートにおける大規模マルチタスク情報抽出データセットの構築. 言語処理学会 第 28 回年次大会, 2022.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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, Volume 1 (Long and Short Papers)**, pp. 4171–4186, 2019.
- [9] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In **Proceedings of the Eighteenth International Conference on Machine Learning**, p. 282, 2001.
- [10] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. Matching the blanks: Distributional similarity for relation learning. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 2895–2905, 2019.