

論述構造解析における事前学習済み言語モデルの有効性検証

内藤 昭一¹ 井之上 直也^{2,3} 乾 健太郎^{2,3}

¹ 株式会社リコー ² 理化学研究所 ³ 東北大学

shohichi.naitoh@jp.ricoh.com {naoya-i,inui}@ecei.tohoku.ac.jp

1 はじめに

論述構造解析 [9] とは、主張や前提といった議論を構成する要素（論述要素）を抽出し、要素間の論述的關係（支持・反対など）を特定するタスクである。論述構造解析はいくつかのサブタスクからなるが、本研究では論述要素への分割は済んでいるものとし、以下の3つのサブタスクに着目する。

論述要素分類:

論述要素を主張や前提などへ分類する

論述関係認識:

論述要素間の論理的關係の有無を分類する

論述関係分類:

論述要素間の關係を支持・反対に分類する

上記サブタスクにおいて重要となるのが、文脈を考慮した論述要素の表現、および論述要素間の關係表現を獲得することである。先行研究 [5, 6, 8] では LSTM ベースのモデルにより文脈を含めた論述要素の表現を獲得することで、高い精度を達成している。

一方、BERT [4] をはじめとして事前学習済み言語モデルを目的タスクへ転移学習する手法が多くタスクで最高精度を更新しており、論述構造解析においても有望な性能が見込まれる。しかし、BERT は一文あるいは文ペアを入力して分類することが想定されており、文脈を含めたうえで論述要素、および要素間の關係を分類する方法は自明ではない。

BERT から文脈内の任意のスパン表現を得る方法はいくつか考えられるが、本研究では初手として關係抽出のタスクで提案された手法 [2] に注目し、論述構造解析において有効性を検証する。文献 [2] では、エンティティをマーカと呼ばれる特殊トークンで囲むことで、文内に存在するエンティティ間の關係表現を得ている。本研究では、論述要素のような文脈内の談話単位においても本手法が有効だと考え、マーカで囲むことにより論述要素、および要素間の關係表現を獲得する。

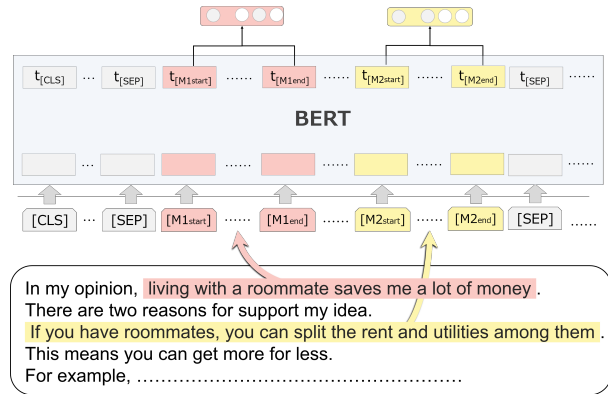


図1: マーカを活用した論述要素のスパン表現、および要素間の關係表現の獲得

実験の結果、本手法は論述要素間の關係を問うサブタスクで特に有効であり、2つのデータセットにおいて最高性能を達成できることが分かった。

2 関連研究

論述構造解析は、ニューラルネットワークを用いた手法により精度が向上している。文献 [8] では、双方向 LSTM をエンコーダとしたポインタネットワークを關係認識に応用し、要素分類とのマルチタスク学習をすることで高い精度を出している。文献 [6] は、接続表現と論述要素を異なる言語単位として扱い、LSTM-Minus によりそれぞれのスパン表現を獲得することで最高性能を達成している。先行研究により、LSTM ベースのモデルから論述要素のスパン表現を得る方法は議論されているが、Transformer ベースのモデルからスパン表現を得る方法は明らかではない。

BERT から文脈を考慮した文表現を得る方法として [3] がある。文献 [3] では、文単位の系列ラベリング問題を解くため、BERT によりエンコードされた $[SEP]$ トークンを文表現として用いている。本研究が対象とする論述要素は一般的に文よりも短い単位で表されるため、 $[SEP]$ トークンよりも粒度の細かいスパン表現が必要と

なる, 文献 [2] では, 関係抽出のタスクで, BERT からエンティティ間の関係表現を得る方法について検証している. 6つの方法を比較した結果, エンティティの前後をマーカと呼ばれる特殊トークンで囲み, 特殊トークンのベクトルを関係表現とする方法が最も良い性能となっている. 本研究のアイデアはこの知見に基づいている.

3 モデル

提案モデルは図1に示すように論述要素をマーカで囲むことで, BERT から文脈を考慮した論述要素のスパン表現, および要素間の関係表現を得る.

3.1 論述要素のスパン表現

論述要素のスパン表現を得る方法について説明する. 論述文を単語系列 $x = \{x_1, \dots, x_T\}$, 論述要素を $a = \{a_1, \dots, a_M\}$ とし, 論述要素をインデックス $a_m = (i, j)$ の範囲とする ($0 \leq i < j \leq T$). ここでは, 単語系列 x とインデックス a_m を入力として, 論述要素のスパン表現を得ることを目的とする.

論述要素の先頭と末尾を表すマーカ $[M_{start}]$ および $[M_{end}]$ を導入し, 論述要素 a_m の前後をマーカで囲んだ単語列 \tilde{x} を BERT への入力とする.

$$\tilde{x} = \{w_1, \dots, [M_{start}], w_i, \dots, w_j, [M_{end}], \dots, w_T\}$$

このとき, 論述要素 a_m のスパン表現 \mathbf{h}_{a_m} は以下により計算する.

$$\mathbf{h}_{a_m} = \text{maxpool}(\mathbf{t}_{[M_{start}]}, \mathbf{t}_{[M_{end}]}) \quad (1)$$

ここで, $\mathbf{t}_{[M_{start}]}$, $\mathbf{t}_{[M_{end}]}$ はマーカ $[M_{start}]$, $[M_{end}]$ に対する BERT 最終層のベクトルである.

論述要素分類では \mathbf{h}_{a_m} を MLP に入力して得られた z_y から, 論述要素 a_m がクラス y に属する確率を計算する.

$$P(y|\tilde{x}) = \frac{\exp(z_y)}{\sum_{y' \in Y} \exp(z_{y'})} \quad (2)$$

3.2 論述要素間の関係表現

論述文内の任意の論述要素間の関係表現を得る方法について説明する. 論述要素 a_m, a_n のインデックスをそれぞれ $a_m = (i, j)$, $a_n = (k, l)$ とする ($0 \leq i < j < k, k < l \leq T$). 単語系列 x とインデックス a_m, a_n を入力とし, 論述要素 a_m から a_n へ方向の関係表現を得ることを目的とする.

関係元を表すマーカ $[M1_{start}]$, $[M1_{end}]$ により a_m を, 関係先を表すマーカ $[M2_{start}]$, $[M2_{end}]$ により a_n

表1: ハイパーパラメータ

| パラメータ | PEC | | MTC | |
|--------------|----------------------|-----------------------|----------------------|-----------------------|
| | BERT _{BASE} | BERT _{LARGE} | BERT _{BASE} | BERT _{LARGE} |
| エポック数 | 10 | 10 | 10 | 10 |
| バッチサイズ | 24 | 4 | 4 | 4 |
| 学習率 | 3e-5 | 1e-5 | 2e-5 | 1e-5 |
| Warmup | 0% | 6% | 10% | 10% |
| Weight decay | 0.0 | 0.0 | 0.1 | 0.1 |

を囲んだ単語列 \tilde{x} を BERT への入力とする.

$$\tilde{x} = \{w_1, \dots, [M1_{start}], w_i, \dots, w_j, [M1_{end}], \dots, [M2_{start}], w_k, \dots, w_l, [M2_{end}], \dots, w_T\}$$

このとき, 論述要素 a_m, a_n の関係表現 $\mathbf{h}_{(a_m, a_n)}$ は, 以下により計算される.

$$\mathbf{h}_{(a_m, a_n)} = [\mathbf{h}_{a_m}; \mathbf{h}_{a_n}] \quad (3)$$

論述関係認識および論述関係分類では, $\mathbf{h}_{(a_m, a_n)}$ を MLP に入力して得られた z_y から, 論述要素 a_m, a_n の関係がクラス y に属する確率を計算する.

$$P_y = \frac{\exp(z_y)}{\sum_{y' \in Y} \exp(z_{y'})} \quad (4)$$

4 実験設定

4.1 データセット

評価のためのデータセットとして, Persuasive Essay コーパス (PEC) [9] と Arg-Microtexts コーパス (MTC) [7] を用いる. PEC はオンラインのエッセイフォーラムから取得した 402 件の論述文からなる. 著者と同様の方法で訓練データと評価データに分割し, ホールドアウト法により評価を行う. MTC は 112 件の短い論述文が集められたデータセットである. 著者が提供するデータ分割に従い, 交差検証法により評価を行う. 両コーパスともに, 訓練データの 10% を開発データとする.

4.2 検証対象

検証する事前学習済み言語モデルとして, BERT_{BASE} と BERT_{LARGE} を対象とする. 実験に使用するパラメータは表1の通りである.

4.3 評価指標

タスクの評価指標として, 先行研究に従い Macro F1 スコアを採用する, また, 総合的な指標として, 全タスクの Macro F1 の平均値を用いる. PEC は 3 つの異なるランダムシードの結果を平均した値, MTC は 10 回の 5 分割交差検証の結果を平均した値を記載する.

表2: 既存手法との比較結果

| | 全体 | 要素分類 | | | | 関係認識 | | | 関係分類 | | | |
|-----------------|---------------------------------|-------------|-------------|------|------|-------|-------------|------|-------|-------------|------|------|
| | | Macro | 大主張 | 主張 | 前提 | Macro | 関係有 | 関係無 | Macro | 賛成 | 反対 | |
| PEC | AC Marked BERT _{BASE} | 81.6 | 81.4 | 86.0 | 68.4 | 90.0 | 80.6 | 68.0 | 93.3 | 82.9 | 97.7 | 68.0 |
| | AC Marked BERT _{LARGE} | 84.4 | 84.8 | 89.0 | 74.0 | 91.3 | 83.6 | 73.0 | 94.2 | 84.9 | 98.0 | 71.7 |
| | LSTM-Minus with ELMo [6] | 81.8 | 85.7 | 91.6 | 73.3 | 92.1 | 80.7 | 67.8 | 93.7 | 79.0 | 96.8 | 61.1 |
| | Joint Pointer Network [8] | - | 84.9 | 89.4 | 73.2 | 92.1 | 76.7 | 60.8 | 92.5 | - | - | - |
| | SVM and ILP [9] | 75.2 | 82.6 | - | - | - | 75.1 | - | - | 68.0 | - | - |
| MTC | AC Marked BERT _{BASE} | 74.6 | 73.1 | - | 55.6 | 90.5 | 69.2 | 49.2 | 89.2 | 81.5 | 86.1 | 76.9 |
| | AC Marked BERT _{LARGE} | 80.6 | 81.9 | - | 70.3 | 93.4 | 75.3 | 59.3 | 91.2 | 84.6 | 88.4 | 80.8 |
| | New Best EG [1] | 78.5 | 87.6 | - | - | - | 72.2 | - | - | 75.7 | - | - |
| | LSTM-Minus with ELMo [6] | 78.2 | 83.5 | - | 72.9 | 94.0 | 73.9 | 57.5 | 90.3 | 77.2 | 84.2 | 70.3 |
| | Joint Pointer Network [8] | - | 81.3 | - | 69.2 | 93.4 | 74.0 | 57.7 | 90.3 | - | - | - |
| SVM and ILP [9] | 76.2 | 85.7 | - | 77.0 | 94.3 | 68.3 | 48.6 | 88.1 | 74.5 | 85.5 | 62.8 | |

表3: 文脈を考慮することの効果

| | 全体 | 要素分類 | 関係認識 | 関係分類 | |
|-----|---------------------------------|------|------|------|------|
| PEC | AC Marked BERT _{BASE} | 81.6 | 81.4 | 80.6 | 82.9 |
| | BERT _{BASE} | 62.0 | 57.3 | 66.4 | 62.4 |
| | AC Marked BERT _{LARGE} | 84.4 | 84.8 | 83.6 | 84.9 |
| | BERT _{LARGE} | 60.4 | 58.0 | 65.0 | 58.2 |
| MTC | AC Marked BERT _{BASE} | 74.6 | 73.1 | 69.2 | 81.5 |
| | BERT _{BASE} | 75.8 | 84.7 | 70.0 | 72.7 |
| | AC Marked BERT _{LARGE} | 80.6 | 81.9 | 75.3 | 84.6 |
| | BERT _{LARGE} | 74.6 | 83.4 | 68.3 | 72.1 |

表4: 提案モデルにおけるマーカの効果

| | 全体 | 要素分類 | 関係認識 | 関係分類 | |
|-----|---------------------------------|------|------|------|------|
| PEC | AC Marked BERT _{BASE} | 81.6 | 81.4 | 80.6 | 82.9 |
| | - マーカ | 78.8 | 80.3 | 77.5 | 78.7 |
| | AC Marked BERT _{LARGE} | 84.4 | 84.8 | 83.6 | 84.9 |
| | - マーカ | 82.2 | 84.1 | 79.4 | 83.0 |
| MTC | AC Marked BERT _{BASE} | 74.6 | 73.1 | 69.2 | 81.5 |
| | - マーカ | 72.7 | 77.9 | 64.0 | 76.3 |
| | AC Marked BERT _{LARGE} | 80.6 | 81.9 | 75.3 | 84.6 |
| | - マーカ | 76.3 | 80.4 | 67.6 | 81.0 |

5 結果

5.1 既存手法との比較

表2に既存手法との比較結果を示す。表中の AC Marked BERT は提案モデルを表している。

PEC, MTC ともに AC Marked BERT_{LARGE} が、総合的に最も良い性能となっている。特に論述要素間の関係を問う関係認識および関係分類のタスクで高い精度を達成している。一方で、論述要素分類では既存手法に劣る結果となった。そのため、提案モデルは単一の論述要素のスパン表現を得るよりも、要素間の関係を表現する用途に適しているといえる。

5.2 文脈を考慮することの効果

文脈を考慮する効果を検証するため、論述要素のみの情報を活用した場合との比較結果を表3に示す。表中の

BERT_{BASE} および BERT_{LARGE} は、論述要素の単語列を入力し、[CLS] トークンのベクトルから予測する一般的な方法による分類結果である。要素分類では単一の論述要素を、関係認識と関係分類では論述要素のペアを入力としている。

表3から、PEC では文脈を考慮することで、一貫して性能が向上していることが分かる。一方、MTC においては PEC と比較して上がり幅が少ないか、もしくは性能が悪化するという結果となった。この原因は論述要素の定義の違いにあると思われる。MTC は “for example”, “because”, “in conclusion” のような接続表現を論述要素に含めているが、PEC では接続表現は論述要素に含まれない。そのため、MTC は PEC と比較し、文脈から得られる有益な情報が少ないと考えられる。

表5: 関係認識における論述要素間の距離による F 値 (関係有) の変化

| | 距離 1 | | 距離 2 | 距離 3 | 距離 4 | 距離 5 以上 |
|---------------------------------|------|------|------|------|------|------------|
| | 明示 | 非明示 | | | | |
| AC Marked BERT _{BASE} | 95.0 | 64.4 | 66.0 | 57.5 | 60.4 | 59.8 |
| - マーカ | 82.2 | 59.9 | 59.6 | 59.0 | 65.2 | 59.7 |
| AC Marked BERT _{LARGE} | 95.2 | 70.6 | 68.6 | 67.2 | 69.9 | 65.2 |
| - マーカ | 87.1 | 62.8 | 62.3 | 61.4 | 66.1 | 63.4 |

5.3 提案モデルにおけるマーカの効果

論述要素をマーカで囲むことによる効果を検証するため、マーカの有無による比較結果を表4に示す。マーカを導入しないケースについては、論述要素の先頭と末尾からスパン表現を得ている。

表3から、ほぼすべてのケースにおいて性能の向上が見られ、マーカで囲むことにより一定の効果があると考えられる。マーカの導入により性能が向上する理由については、今後、勾配やアテンションの値を比較することで詳細な分析をしたい。

5.4 分析

論述構造解析において、関係認識は高度な推論を要する難しいサブタスクである。関係認識における提案モデルの課題を探るため、比較的データ量の多い PEC を対象に分析を行う。

表5は論述要素間の距離による F 値の変化である。距離 1 は、論述要素が隣接していることを意味する。また、隣接する論述要素については接続表現が存在する明示的關係*1と存在しない非明示的關係に分けている。

表5から、明示的な関係についてはかなり高い精度で解けていることが分かる。また、提案モデルに関しては、BERT のサイズによって大きな差はなく、精度向上の余地は少ないと考えられる。

一方、非明示的な関係、および論述要素が隣接していない関係については、F 値が 60 から 70 ポイント程度に留まっている。これらの関係を正しく予測するためには、内容についての深い理解が求められるが、現状の手法では難しい問題だと言える。しかし、BERT のサイズを大きくすることで精度の向上が見られることから、関係認識において内容への着目がより強まっている可能性がある。

*1接続表現とは“therefore”, “because”, “as a result”など談話単位をつなぐ表現のことであり。[9]を参考に定義している。

6 おわりに

本研究では事前学習済み言語モデルから文脈を考慮した論述要素のスパン表現、および要素間の関係表現を得る方法を提案し、論述構造解析における有効性を検証した。実験の結果、論述要素間の関係を問うサブタスクにおいて特に有効であり、全体として既存手法を上回る性能を達成できることを確認した。

また、分析の結果、最も性能が良いモデルであっても、論述要素間に接続表現がない非明示的な関係や隣接していない関係の認識は十分にできていないことが分かった。このような関係を正しく認識させることは今後の課題である。

参考文献

- [1] S. Afantenos et al. “Comparing decoding mechanisms for parsing argumentative structures”. In: *Argument and Computation* 9.3 (2018), pp. 177–192.
- [2] L. Baldini Soares et al. “Matching the Blanks: Distributional Similarity for Relation Learning”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019, pp. 2895–2905.
- [3] A. Cohan et al. “Pretrained Language Models for Sequential Sentence Classification”. 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)*. Association for Computational Linguistics, 2019, pp. 3691–3697.
- [4] J. 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, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 2019, pp. 4171–4186.
- [5] S. Eger et al. “Neural End-to-End Learning for Computational Argumentation Mining”. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2017, pp. 11–22.
- [6] T. Kuribayashi et al. “An Empirical Study of Span Representations in Argumentation Structure Parsing”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019, pp. 4691–4698.
- [7] A. Peldszus. “Towards segment-based recognition of argumentation structure in short texts”. In: *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, 2014, pp. 88–97.
- [8] P. Potash et al. “Here’s My Point: Joint Pointer Architecture for Argument Mining”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2017, pp. 1364–1373.
- [9] C. Stab and I. Gurevych. “Parsing Argumentation Structures in Persuasive Essays”. In: *Computational Linguistics* 43.3 (2017), pp. 619–659.