

BERTとRefinementネットワークによる 統合的照応・共参照解析

植田 暢大[†] 河原 大輔[†] 黒橋 禎夫^{†‡}

[†]京都大学 [‡]科学技術振興機構 CREST

{ueda,dk,kuro}@nlp.ist.i.kyoto-u.ac.jp

1 はじめに

自然言語で書かれた一貫性のあるテキストの背後にはエンティティ間の述語項構造関係、共参照関係、橋渡し照応関係などがあり、これらを明らかにすることは計算機が自然言語を理解する上で不可欠である。

述語項構造関係とは、述語を中心とするその述語の「誰が」や「何を」に相当する項からなる関係であり、その関係を明らかにするタスクは述語項構造解析とよばれる。述語項構造解析は、対象の項と述語の間に係り受けの関係がある場合に特に格解析とよばれ、ない場合にゼロ照応解析とよばれる。日本語の格解析は90%ほどの精度で解けるのに対し、ゼロ照応解析の精度は、係り受けなどの手がかりが使えないため50%台に留まっている [1, 2, 3, 4]。

橋渡し照応関係とは、名詞間に成り立つ関係であり、例えば「向こうに家が見えた。屋根は白かった。」という文において「家」と「屋根」は「家の屋根」という関係にあり、「屋根」に対してこの「家」は橋渡し照応関係にあるという。この関係を解析する橋渡し照応解析も非常に難しいタスクであり、Sasanoら [5] の先行研究では F1 スコア 42.2 に留まっている。

本研究では、これら関係の統合的解析に BERT [6] を適用する。BERT は、Transformer [7] をベースにしたモデルで、様々な自然言語処理タスクにおいて最高性能を達成している。BERT を用いた結果、入力文に関する素性を一切用いることなく、既存手法を大きく上回る結果が得られた。

従来、精度向上のために述語や格の間の関係性を明示的に捉えるようなモデルが成功を収めていた [2, 8]。一方で、BERT を使用すると個々の予測が独立になってしまう。そこで、一度 BERT で予測を行った後、その結果に基づいてよりもっともらしい結果を得る Refinement ネットワークを導入し、さらなる精度の向上を確認した。

2 関連研究

日本語の述語項構造解析では、近年他のタスクとの同時学習が効果的であることが示されている [4, 3]。Shibata ら [4] は、述語項構造解析と共参照解析を同時学習することで述語項構造解析の精度が向上することを示した。Omori and Komachi [3] は、述語だけでなく、事態性名詞とよばれる項を取る名詞に対しても項構造解析を行うことで、述語項構造解析の精度を改善した。しかし、いずれの研究も共参照解析と事態性名詞の項構造解析の両方を行ってはいない。本研究ではこれら両方の同時学習に加え、名詞に対する橋渡し照応解析も同時に扱う。

述語項構造解析のアプローチの一つとして、複数の述語や項の間の関係を考慮するということが考えられる。Ouchi ら [8] は Grid-RNN という機構を導入することで複数の述語間の関係を捉えることを試み、その有効性を示した。Matsubayashi and Inui [2] は、RNN の代わりに attention 機構を採用することで、語順や距離に依存しない述語間関係を捉えることを試み、Ouchi らに対する優位性を示した。

英語では日本語の述語項構造解析に類似するタスクとして Semantic Role Labeling (SRL) がある。Iyu ら [9] は SRL において、一度システムが予測した結果を使用して、よりもっともらしい結果を出力する Refinement ネットワークを導入した。本研究でも同様の考えに基づき、システム予測を洗練するようなネットワークを導入する。

Kurita ら [1] は、日本語の述語項構造タグ付きコーパス不足の問題に着目し、敵対的学習を用いて生コーパスを活用する手法を提案した。

以上の関連研究はすべて、双方向 RNN を使用しているのに対し、本研究では self-attention 機構に基づいた BERT を使用した。

橋渡し照応解析についての先行研究は多くない。Sasano ら [5] は大規模コーパスから「A の B」のよう

な表現を集めて辞書を構築し、この辞書に基づく解析手法を提案している。我々の知る限りニューラルネットワークに基づく橋渡し照応解析の既存研究はない。

3 提案手法

3.1 ベースモデル

実験に使用した BERT に基づくベースモデルを図 1 左側に示す。図はモデルが、述語「投函した」の格の項が「[著者]」となるスコアを計算する場合を表す。

3.1.1 入力表現

入力文書に対して文区切りと形態素分割をした後、BPE [10] を用いてサブワードに分割する。文区切りと形態素分割に関してはコーパスに正解が付与されているので、それを用いる。Devlin ら [6] に倣い、入力文書の先頭と末尾にそれぞれ [CLS] と [SEP] トークンを挿入し、BERT モデルに入力する。この際、5つの特殊トークン [著者]・[読者]・[不特定:人]・[NULL]・[NA] を入力系列の末尾に挿入する。[著者]・[読者]・[不特定:人] はそれぞれ外界の照応先に対応するトークンであり、[NULL]・[NA] はそれぞれ、述語項構造解析において項が存在しない場合、共参照解析において共参照先が存在しない場合を表す。BERT の pre-training は系列長の最大を 128 とした。入力系列が 128 を超える文書は学習・評価から除外されるが、今回使用したコーパス内にそのような文書は存在しなかった。

3.1.2 出力層

予測は Shibata ら [4] や Kurita ら [1] と同様、項選択モデルに基づく。BERT 最終層のベクトルを MLP に入力し、文書中の全サブワード対に対してそれぞれの格のスコアを計算し、各述語と格に対し、スコア最大の項をその述語の項に採用する。つまり、述語サブワード p_i に対し、項候補サブワード a_j が格 c の項になる確率は、

$$P(a_j|p_i, c) = \frac{\exp(s_c(a_j, p_i))}{\sum_k \exp(s_c(a_k, p_i))} \quad (1)$$

$$s_c(a_j, p_i) = \mathbf{v}^T \tanh(W_c \mathbf{a}_j + U_c \mathbf{p}_i) \quad (2)$$

と表される。ここで、 \mathbf{p}_i , \mathbf{a}_j はサブワード p_i , a_j に対応する BERT 最終層のベクトルを表す。

なお、コーパスのアノテーションは基本句という 1つの内容語と 0 個以上の付属語からなる単位に基づいている一方、モデルによる予測はサブワード単位である。ここでは、基本句中の内容語形態素の先頭サブワードを、その基本句を代表するサブワードとして学習及び予測に用いる。

3.2 同時学習

本研究では、述語項構造解析に加え、共参照解析、橋渡し照応解析も同時に行う。これらの解析は互いに関係の深いタスクであるので、同時学習を行うことで互いに利得があると考えられる。これら解析はいずれも項選択の枠組みで行われ、共参照解析については述語について格の項を選択すると同様にエンティティを共有するメンションを選択する。橋渡し照応解析については、文書中の体言に対して橋渡し照応関係にある体言を選択する。

本実験で用いたコーパスには述語だけでなく、サ変名詞などの名詞に対しても格関係がアノテーションされている。このような述語と同様に項を取る名詞を事態性名詞とよぶ。本研究ではこれら事態性名詞も項構造解析の対象とする。

3.3 格間関係の考慮

図 1 左側に示すベースモデルは、BERT の self-attention 機構によって他の単語の情報を共有しているとはいえ、それぞれの述語と格について項の予測は独立になっている。すなわち格間の整合性を明示的には考慮していない。この問題に対処するため、1 度予測した結果を元に、その結果を修正するように再度予測を行う Refinement ネットワークを導入する。図 1 右側に示すように、この Refinement モデルでは、ベースモデルで予測した確率を出力層の隠れベクトルにそれぞれの格ごとに結合する。

ベースモデルによる予測を $P(a_j|p_i, c)$ とすると、Refinement モデルによる予測は以下で計算される。

$$P'(a_j|p_i, c) = \frac{\exp(s'_c(a_j, p_i))}{\sum_k \exp(s'_c(a_k, p_i))} \quad (3)$$

$$s'_c(a_j, p_i) = \mathbf{v}'^T \tanh(\mathbf{h}_{i,j,c}) \quad (4)$$

$$\mathbf{h}_{i,j,c} = (W'_c \mathbf{t}'_j + U'_c \mathbf{t}'_i) \circ \mathbf{P}(a_j|p_i, c) \quad (5)$$

ここで、 \mathbf{p}'_i , \mathbf{a}'_j は p_i , a_j に対応する Refinement ネットワーク側の BERT 最終層のベクトルである。 \circ はベクトルの結合を表す。 $\mathbf{P}(a_j|p_i, c)$ はベースモデルの結果の確率を全ての格に渡って結合したもので、

$$\mathbf{P}(a_j|p_i, c) = P(a_j|p_i, c_0) \circ \dots \circ P(a_j|p_i, c_K) \quad (6)$$

と表される。なお、 K は解析対象の格の数を表す。

4 実験

4.1 実験設定

同時学習の効果を確かめるため、ベースモデルにおいて述語項構造解析のみを行った場合と、共参照解析や、

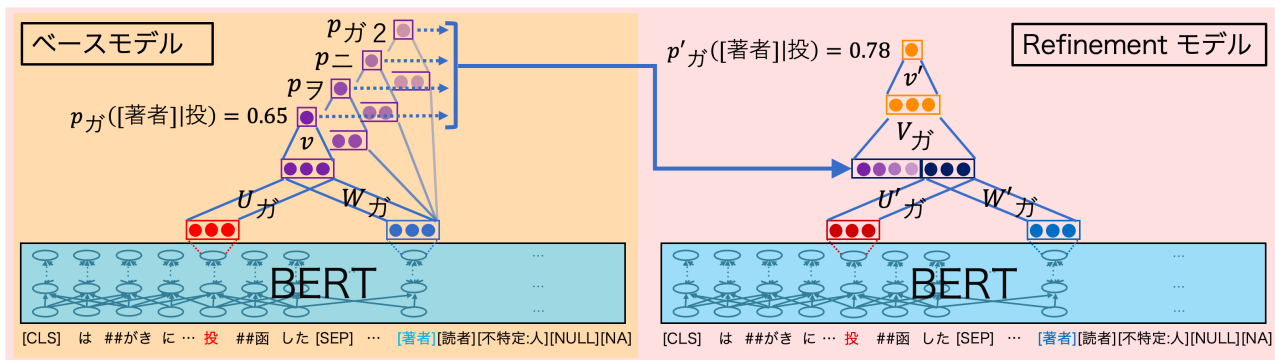


図 1: 提案モデルのネットワーク

事態性名詞に対する項構造解析・橋渡し照応解析も含めた場合の実験を行った。また、全てを含めた場合に、Refinement ネットワークを導入した場合の実験を行った。BERT モデルには日本語 Wikipedia 1,800 万文で 25 エポック pre-training された BERT_{LARGE} モデルを使用した。

学習及び、評価には京都大学ウェブ文書リードコーパス (KWDLIC) [11] を使用した。KWDLIC は、ウェブから収集した約 5,000 文書、15,000 文に対して述語項構造や照応関係がアノテーションされている。

述語項構造解析を行う場合は、ガ格・ヲ格・ニ格・ガ2格の4つの格を対象に、学習時は用言および事態性名詞について解析し、評価は用言に対する結果のみを対象とした。橋渡し照応解析を行う場合は、体言に対してそのノ格を解析した。共参照解析は体言に対して行った。Shibata ら [4] に倣い、外界照応として著者・読者・不特定:人を考え、いずれの解析についても正解の基本句と同じ共参照クラスタ内の基本句は全て正解とした。

学習はクロスエントロピーロスを用いて、4 エポック行った。また、各実験は異なる初期値で、5 回行った。

Refinement モデルに関して、損失はベースモデルの予測と Refinement ネットワークの予測それぞれのクロスエントロピーロスの和を取った。また、Refinement ネットワークの損失は、Refinement ネットワークのみを学習させるため、ベースモデル側には伝搬させないようにした。

4.2 結果

F1 スコアで評価した実験の結果を表 1 に示す。「Base」はベースモデルを表し、「+coref」、「+noun」、「+bridge」はそれぞれ同時に共参照解析、事態性名詞の項構造解析、橋渡し照応解析を行った場合を表す。Refine は 3.3 節で述べた格間関係を考慮したモデルである。また、括弧内に結果の 95% 信頼区間を示す。

Shibata ら [4] の先行研究と比べ、BERT を用いることでゼロ照応解析の精度が大きく向上する結果となった。また、BERT は共参照解析にも有効で、先行研究の共参照解析の結果を 2.6 ポイント上回るだけでなく、同時学習によりゼロ照応解析の精度も向上した。その他、事態性名詞に対する解析もゼロ照応解析の精度を有意に向上させることが分かった。橋渡し照応解析に関しては有意差は見られなかった。さらに、Refinement ネットワークを使用することで、一貫して精度が向上することが確かめられた。橋渡し照応解析の精度については解析対象が異なるので Sasano らの先行研究の F 値 42.2 と直接の比較はできないが、十分にリーズナブルなスコアであると考えられる。一方で、格解析は先行研究を下回る結果となっているが、これは提案手法が係り受けや選択選好の素性を一切使用していないことに起因すると考えられる。

4.3 議論

Refinement モデルが実際に格間関係を考慮しているのかを確かめるため refine 前後の解析例を取り上げる。これは全ての同時学習を行った場合の結果であり、下線部の述語について予測や正解を示している。

- (1) クロノグラフ機能もついて高級感とスポーティさを兼ね備えたモデル。

(refine 前)	ガ格: モデル	ニ格: モデル
(refine 後)	ガ格: 機能	ニ格: モデル
(正解)	ガ格: 機能	ニ格: モデル

- (2) ボルタに仕事を与える女の子。

(refine 前)	ガ格: NULL	coref: 著者
(refine 後)	ガ格: 著者	coref: 著者
(正解)	ガ格: 不特定:人	coref: NA

例 (1) について、refine 前ではガ格とニ格に同じ項を選んでしまっているのに対し、refine 後ではどちらも

	格解析	ゼロ照応解析	共参照解析	橋渡し照応解析
Base	85.03 (± 0.60)	64.03 (± 0.18)	-	-
Base+coref	84.66 (± 0.24)	64.50 (± 0.55)	71.01 (± 0.30)	-
Base+coref+noun	84.99 (± 0.32)	65.32 (± 0.75)	71.52 (± 0.09)	-
Base+coref+noun+bridge	84.85 (± 0.40)	65.28 (± 0.49)	71.52 (± 0.49)	47.08(± 1.12)
Refine+coref+noun+bridge	85.98 (± 0.62)	65.61 (± 0.34)	72.09 (± 0.48)	49.58 (± 0.89)
Shibata ら [4]	89.2	58.1	68.5	-

表 1: 述語項構造解析・共参照解析・橋渡し照応解析の F1 スコア

正しく項を選択できている。これは、Refinement ネットワークにおいて、「モデル」が「ついて」のガ格であるスコアを計算する際に、「モデル」が「ついて」のニ格であるということ considering、スコアを下げているためと考えられる。例 (2) では、洗練によって誤りではあるが、ガ格の項が著者と判断されるようになっている。日本語では、「A が B だ。」という文において A と B が共参照関係にあるような場合が頻繁に存在するが、例 (2) は、refine 前において「女の子」が「著者」と共参照関係にあると判断されていることを考慮した結果と考えられる。

Refinement モデルの解析誤りの他の例を示す。

- (3) また、サブディスプレイもより大型の多色の 1.2 インチ 6 万 5 5 3 6 色のものを搭載。
 (システム) ガ格: 色
 (正解) ガ格: ディスプレイ
- (4) 蕎麦を打つときに手や板や麺棒に付かないようにまぶす粉です。
 (システム) ガ格: 粉 ニ格: 棒
 (正解) ガ格: 蕎麦 ニ格: 板 手棒

例 (3) はシステムが「大型の」のガ格を「色」と判断しており、選択選好が十分に捉えられていないことが分かる。これは、Shibata ら [4] の先行研究に比べ、格解析の精度が低くなっている一因と考えられる。例 (4) は、ニ格は正解しているものの、ガ格が「粉」と誤って予測している例である。この例は、「粉をまぶす」という出来事と「粉が付かない」という出来事が共起しないという常識を考慮できれば解ける可能性がある。

5 おわりに

本研究では、BERT を用いて述語項構造解析・共参照解析・橋渡し照応解析の同時学習を提案し、有効性を示した。さらに、Refinement ネットワークを使用することで、BERT の予測が独立になってしまうという問題に対処した。

今後は、data augmentation による格解析も含めた精度向上を試みるほか、スクリプト知識などの常識も考慮することが可能なモデルを目指したい。

謝辞

本研究は科学技術振興機構 CREST「知識に基づく構造的言語処理の確立と知識インフラの構築」(JP-MJCR1301) の支援のもとで行われた。

参考文献

- [1] Shuhei Kurita, Daisuke Kawahara, and Sadao Kurohashi. Neural adversarial training for semi-supervised Japanese predicate-argument structure analysis. In *Proceedings of ACL*, pp. 474–484, 2018.
- [2] Yuichiroh Matsubayashi and Kentaro Inui. Distance-free modeling of multi-predicate interactions in end-to-end Japanese predicate-argument structure analysis. In *COLING2018*, pp. 94–106, 2018.
- [3] Hikaru Omori and Mamoru Komachi. Multi-task learning for Japanese predicate argument structure analysis. In *Proceedings of NAACL*, pp. 3404–3414, 2019.
- [4] Tomohide Shibata and Sadao Kurohashi. Entity-centric joint modeling of Japanese coreference resolution and predicate argument structure analysis. In *Proceedings of ACL*, pp. 579–589, 2018.
- [5] 笹野遼平, 河原大輔, 黒橋禎夫. 自動獲得した知識に基づく統合的な照応解析. 言語処理学会 第 12 回年次大会, pp. 480–483, 横浜, 2006.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL*, pp. 4171–4186, 2019.
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS2017*, pp. 5998–6008, 2017.
- [8] Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. Neural modeling of multi-predicate interactions for Japanese predicate argument structure analysis. In *Proceedings of ACL*, pp. 1591–1600, 2017.
- [9] Chunchuan Lyu, Shay B. Cohen, and Ivan Titov. Semantic role labeling with iterative structure refinement. In *Proceedings of EMNLP-IJCNLP*, pp. 1071–1082, 2019.
- [10] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *ACL2016*, pp. 1715–1725, 2016.
- [11] Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. Building a diverse document leads corpus annotated with semantic relations. In *Proceedings of PACLIC*, pp. 535–544, 2012.