

前方文脈の埋め込みを利用した日本語述語項構造解析

今野 颯人¹ 松林 優一郎^{2,1} 大内 啓樹^{2,1} 清野 舜^{1,2} 乾 健太郎^{1,2}

¹ 東北大学 ² 理化学研究所 AIP センター

{ryuto,kiyono,inui}@ecei.tohoku.ac.jp

{yuichiro.matsubayashi,hiroki.ouchi}@riken.jp

1 はじめに

述語項構造とは、文章内の述語とその項の間の関係を規定する構造であり、事象や行為を表す単位として用いられる。例えば次の文、

- (1) [彼女] は [パン] を 食べた。

では、「食べた」という表現が述語であり、「彼女」や「パン」という表現が述語「食べた」の項である。述語項構造を解析することで、複雑な構造を持った文においても、文の理解に重要な「誰が、何を、どうした」といった構造化された意味関係を獲得することができる。

日本語では、述語と意味的關係を持つ項が直接の係り関係にない事例（ゼロ照応）がたびたび現れる。例えば、

- (2) (ϕ ガ) 風邪を ひいた ので、[私] は学校を 休んだ。

では、「ひいた」と「休んだ」が述語であり、どちらも「私」が項（主格）であるが、「ひいた」と「私」のペアは直接の係り関係にないため、ゼロ照応の事例に分類される。ゼロ照応の解析精度は、解析対象を述語と同一文内にある項に限っても F_1 値で 58% 程度 [1] と低い水準に留まっており、ゼロ照応解析の精度向上は日本語の述語項構造解析における主要な課題となっている。

こうした背景から、本稿では日本語のゼロ照応解析の精度向上を目的として、項構造解析モデルに前方文脈の情報を取り込む新しいモデルを提案する。日本語の解析において、前方文脈の情報を取り入れることは項の推定に直感的に有効である。例えば、同一文書内に連続して出現する次の 2 文、

- (3) なにものかが大量にコピーした偽造用紙を、本物の投票用紙と偽って候補 [者_{ni}] に 売り付け しようとしていたとみて調べている。
- (4) ... 男性の声で電話がかかり、候補 [者_{ni}] 自身が出ると「投票用紙を一枚五、六万円で買わないか」と 持ち掛け、候補者以外が電話に出ると、

すぐに切れるという不審な電話があったという。
(NAIST Text Corpus: 950113-0145-950113207.ntc)

では、例文 (4) で「持ち掛ける」の与格（二格）が「候補者」となっている。しかし、この項は述語と直接的な係り受け関係になく、複雑な構文構造や「電話をかけてきた人が電話に出た人に何かの用事がある」といった常識的知識を理解しなければ導けない、解析難易度の高い事例である。一方で、その直前の文である例文 (3) には、「なにものかが」「偽造用紙を」「候補者に」「売りつける」といった意味関係が、述語と項の直接的な係り受け関係を持って明瞭に書かれている。このような場合、一般に前方文脈で文の主役となっているエンティティは後方の文でも項として言及されることが多い。また、前文で示される直接的な意味関係を用いて「候補者に売りつける」ために「候補者に持ち掛ける」といった意味的なつながりをとらえることで、例文 (4) のような複雑なケースの解析を容易にできる可能性がある。したがって、前方文脈の意味構造を何らかの形で埋め込み、後方の文の解析に利用すれば、ゼロ照応のような統語的な手がかりの少ない事例の解析に貢献すると期待できる。

そこで本稿では、多層双方向リカレントニューラルネット (RNN) を用いた既存の日本語述語項構造解析器 [1] を拡張し、入力として解析対象の文に加えて前方の n 文を受け取って解析を行うモデルを構築する。実験ではこのモデルを用いて、日本語述語項構造解析において前方文脈を解析に用いる効果を検証する。

結果として、前方の文を入力に結合するシンプルなモデルにおいても、適切な長さで前方の文の情報を取り入れることで、ゼロ照応のような統語的な手がかりの少ない事例の解析精度を向上させることが分かった。

2 関連研究

日本語述語項構造解析の直近の研究では、多層双方向 RNN を用いた end-to-end の解析モデルが複数提案され、成功を収めている [1-3]。しかし、これらのモデルは

解析対象の述語が含まれる 1 文の情報のみから解析を行うため、文を超えた文脈情報を捉えることはできない。

前方文脈を利用するニューラル解析モデルとして、Shibata ら [4] の研究が挙げられる。彼らは文書内のそれぞれの文に対して、逐次的に共参照解析と述語述語項構造解析を解析し、前方の文における双方の結果を利用し、エンティティごとに用意したベクトルに文脈を埋め込んでいくことで、これら両方のタスクの解析精度が向上することを示した。Shibata らのモデルは、前方の文の共参照解析や項構造解析を陽に解き、エンティティベクトルを更新することで文脈が考慮される。一方で我々の方法では、これらの関係をモデルが陽に解けるかわからず、前方の n 文のすべての情報を RNN でエンコードする形で利用し、前方文の情報が解析に総合的に有益かどうかを検証する。

機械翻訳の近年の研究でも文脈情報を用いる手法が模索されてきている [5–7]。例えば Tiedeman ら [5] は、原言語文および目的言語文の系列にそれぞれ前文を連結してニューラル機械翻訳モデルの学習を行うことで、翻訳精度が向上することを報告した。

3 タスク設定

本研究では、述語項構造解析の研究で利用される主要なコーパスの一つである、NAIST Text Corpus (NTC) 1.5 版 [8] の注釈の仕様に基づき実験を行う。既存研究 [1, 2] に習い、述語項関係の中でも特に述語と項が同一文内にあるものを解析対象とする。この設定は、文内の情報だけから項を特定するには複雑な関係を紐解く必要が生じる例文 (4) のような例において、前方文脈の情報を得ることで解析の難易度を緩和できるかどうかを検証するためのものである。また、項が述語と同一文節内にある場合は評価の対象外とする。

本稿のモデルは、各述語に対し、その述語がある文中からガ、ヲ、ニ格に対応する項をそれぞれ高々一つ出力する。項はその主辞となる 1 単語を選ぶ形で出力される。評価の際は、システムが出力した語が NTC で正解とされる語のいずれかと一致した場合に正解とする。

4 提案モデル

我々のモデルは、Matsubayashi & Inui [1] でベースラインとして用いられた多層双方向 RNN による end-to-end 解析モデル (SELFATT モデル) を、入力に複数文を取るモデルに拡張したものである。我々のモデルのネットワークを図 1 に示す。

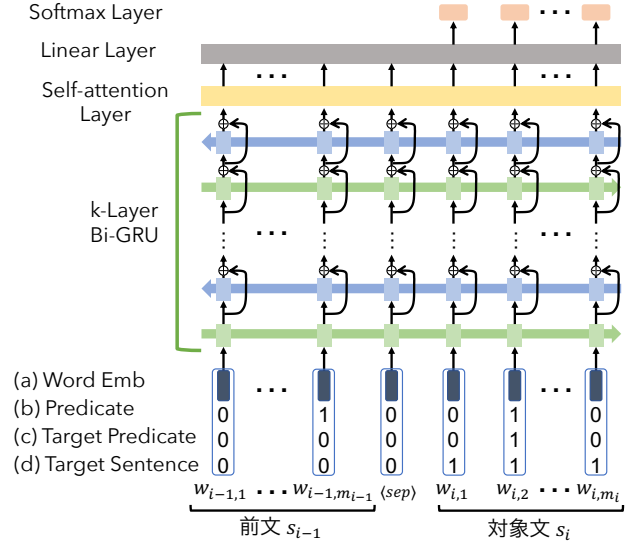


図1: 提案モデルのネットワーク

今、解析対象の述語を持つ文（解析対象文）を文章中の i 番目の文とし、 $s_i = w_{i,1}, \dots, w_{i,m_i}$ で表す（ただし、 $w_{i,j}$ は文 s_i の j 番目の単語、 m_i は文 s_i の単語数）。我々の解析モデルは、解析対象の文に加えて前方の n 文 s_{i-1}, \dots, s_{i-n} を入力として取る。

具体的な入力として、モデルは (a) 文 s_{i-n} から s_i までの単語系列の、各語の埋め込み表現、(b) 系列中出现する全ての述語の位置を表すバイナリ表現、(c) 解析対象とする述語の位置を表すバイナリ表現、(d) 系列における解析対象文の範囲を表すバイナリ表現を受け取る。これら (a) (b) (c) (d) を各単語ごとに一つのベクトルに結合してモデルに入力する。ここで、(d) のバイナリ列は、入力される複数文の中で解析対象文がどの部分かをモデルに教えるために我々が新たに追加したものである。これに加えて、文の境界には $\langle sep \rangle$ という特殊トークンを挿入する。

モデルは入力されたベクトルから多層双方向 RNN を用いてラベルを推定する。この多層双方向 RNN は、奇数層を順方向、偶数層を逆方向に計算し、結果を次の層に入力する。この際、各層では、直前の RNN 層の出力と現在の RNN 層の出力の和を次の層に渡す residual connection [9] を採用している。RNN には GRU [10] を採用する。その後、多層双方向 RNN の最終出力に、機械翻訳の Transformer モデル [11] で利用された multi-head self-attention 層を適用し、続いて線形層で 4 次元ベクトルに変換した後、softmax 層で各単語におけるガ、ヲ、ニ、NONE の各ラベルの出力確率を計算する。ラベ

表1: NTC 1.5 における項の数

		ガ格	ヲ格	ニ格
訓練	DEP	36,877	24,624	5,741
	ZERO	12,201	2,130	464
開発	DEP	7,414	5,044	1,611
	ZERO	2,660	443	138
評価	DEP	13,982	9,395	2,488
	ZERO	4,990	903	260

ルの教師信号は、解析対象文の単語にのみ与える。^{*1}

各単語に対する最終的な項の出力は、解析対象文の中から最尤の1語を選択することで決定される。まず、解析対象文の各単語について、ガ、ヲ、ニ、NONEの中から最も確率の高いラベルを選出し、その単語が持つラベルの候補とする。その後、NONEを除くガ、ヲ、ニの各ラベルについて、同じラベルを持つ候補のうちから、文中で最大の確率を持つ単語をその格の項として一つ選び出力する。文中に当該ラベルを持つ候補が一つも選出されない場合は、その格は出力しない。

5 実験

実験では、提案モデルで入力として用いる前方の文の数を変えることにより、日本語述語項構造解析において前方文脈を考慮することの効果を検証する。具体的には、入力として解析対象文のみ、解析対象文 + その前方1文、解析対象文 + その前方2文を用いる3パターンのモデルを比較する。また、前方の文を結合することによる系列長の増加に対して、self-attention層がどのように働くかを観察するため、self-attention層を省略したモデルとも比較を行う。

5.1 実験設定

実験データである NTC 1.5 は既存研究の分割に従い、訓練セット、開発セット、評価セットに分割する [12]。表 1 に各セットの項の数を示す。表中の DEP は直接係り受けあり、ZERO はゼロ照応の事例を表している。

評価は、適合率、再現率、 F_1 値を求めることで行う。実験では、各モデルについて、ランダムに設定した seed 値を用いて同一の学習データとハイパーパラメータで3回学習を行い、それらの評価値の平均を表に記載した。単語埋め込みベクトルの次元数は 256、多層双方向 GRU の層の深さ k は 8、GRU の隠れ状態の次元数は 256、GRU の dropout 率を 0.1 に設定した。学習では、開発セットでのモデルの F_1 値が5回改善されなかった場合に学習を終了し、開発セット上で F_1 値が最大とな

るモデルを評価に用いた。Loss 関数には交差エントロピー、最適化手法には Adam [14] を採用した。各モデルの学習係数の値は、0.0001, 0.00005, 0.00002 で訓練し、 F_1 値が最大となるものを採用した。単語埋め込み表現の初期値は、Matsubayashi & Inui (2017) [13] で用いられたものと同じの学習手法を適用し、日本語 Wikipedia 2016 年 9 月 1 日のダンプデータより得られた単語埋め込み表現を使用した。これらの単語埋め込み表現は各語を見出し語に変換して学習、利用した。NTC 上の単語が Wikipedia 上の未知語となる場合には、予め Wikipedia 上で頻度 10 回未満の単語を置き換えて学習した未知語用の埋め込み表現に置き換えた。これらの単語埋め込み表現は学習時に更新した。

5.2 実験結果

評価セットにおける各モデルの解析性能を表 2 に示す。まず、self-attention 層を省いた BASE モデルについて見ると、解析対象文に前方の文を繋げて入力するモデルは、単一文を入力するモデルと比べて、全体の F_1 値では大きな効果は見られなかった。一方で、ゼロ照応の事例では、前方1文を追加するモデルは単一文を入力するモデルと比べて F_1 値で 0.5 ポイントの向上が見られた。前方2文を加えるモデルは、ゼロ照応事例の解析精度が他のモデルに比べ相対的に低く、前1文を加えるモデルから 1.2 ポイントの F_1 値低下となった。

Self-attention 層を用いた ATTN モデルについて見ると、まず、我々の実装において ATTN モデルは BASE モデルと比較して全体の F_1 値ではやや低い性能となった。一方で、ゼロ照応の事例では、対応する BASE モデルと比べて精度の向上が認められた。特に前方1文を追加するモデルでは BASE + 前1文モデルに比べて1ポイントの F_1 値向上がみられ、self-attention 層を追加することがゼロ照応の解析精度向上に貢献することが分かった。また、ATTN モデルでは、前方1文を加えるモデルが全体の F_1 値で最も良い性能を示し、単一文を入力する場合に比べて 0.4 ポイント向上した。入力する前方の文を2文にした場合では、BASE モデルの場合と同様に、ゼロ照応事例の解析精度低下が見られた。

これらの結果から、前方文脈の利用はゼロ照応の事例の解析に有効であるものの、前方の文を結合する単純なモデルでは直接係り受けを持つ事例の精度低下を招くほか、入力系列を必要以上に長くすることはかえって精度の低下を招くことが分かった。Self-attention 層を加えたモデルでは、前方1文を追加するモデルが最も高い性能を示し、また、そのゼロ照応解析の精度は、比較した

^{*1} 予備実験で、前方の文の語にもラベルの教師信号を与える場合も検証したが、解析対象文にのみ信号を与えるほうが良い性能を示した。

表2: NTC 1.5 評価セットにおける評価結果. BASE は self-attention を省いたモデル. ATTN は self-attention を用いたフルモデル.

Model	ALL			DEP				ZERO			
	F_1	適合率	再現率	ALL	ガ格	ヲ格	ニ格	ALL	ガ格	ヲ格	ニ格
BASE	80.18	85.43	75.55	87.59	88.03	93.42	62.61	47.20	50.10	39.6	12
BASE + 前 1 文	80.23	84.93	76.02	87.72	88.10	93.16	64.57	47.72	50.83	38.3	13
BASE + 前 2 文	80.05	85.09	75.62	87.93	88.03	93.18	65.51	46.57	49.79	37.8	12
ATTN	79.52	83.11	76.22	87.06	87.52	92.90	62.95	47.77	50.84	38.4	15
ATTN + 前 1 文	79.96	83.67	76.58	87.53	87.85	92.85	65.44	48.73	51.68	39.0	18
ATTN + 前 2 文	79.96	85.11	75.41	87.59	87.73	93.05	64.77	47.18	50.08	37.9	17

表3: NTC 1.5 評価データセットにおける既存モデルとの比較

	ALL			DEP				ZERO			
	F_1	適合率	再現率	ALL	ガ格	ヲ格	ニ格	ALL	ガ格	ヲ格	ニ格
BASE + 前 1 文	80.23	84.93	76.02	87.72	88.10	93.16	64.57	47.72	50.83	38.3	13
Ouchi+2017-sgl [2]	81.15	—	—	88.10	88.32	93.89	65.91	46.10	49.51	35.1	9.8
Ouchi+2017-multi [2]	81.42	—	—	88.17	88.75	93.68	64.38	47.12	50.65	32.4	7.5
M&I 2017 [13]	83.50	—	—	89.89	91.19	95.18	61.90	51.79	54.69	41.8	17
M&I 2018 [1]	83.94	86.58	81.46	90.26	90.88	94.99	67.57	55.55	57.99	48.9	23

モデル全体を通して最も高い数値を示したが、事例全体の精度では BASE モデルに劣っており、直接係り受けの事例とゼロ照応の事例の精度を同時に向上する機構を検討することが今後の課題である。

表 3 には、日本語述語項構造解析における既存モデルの精度を示した。我々のモデルは、Matsubayashi & Inui [1] でベースラインとして用いられた SELFATT モデルを拡張したものであったが、実装の詳細の違いにより、報告されている結果より全体的に低い精度にとどまった。今後、既存研究と実装をそろえた実験設定で比較を行っていく予定である。

6 おわりに

本稿では、多層双方向 RNN を用いた既存の日本語述語項構造解析モデルを拡張し、入力に前方の n 文を加えて解析を行うモデルを構築した。結果として、ゼロ照応の事例においては、前方 1 文を追加したモデルで精度の向上がみられた。また、self-attention 層を加えたモデルがゼロ照応の事例で総じて高い性能を示した。一方で、事例全体の解析精度は単一文を入力するモデルと比べて有意な性能向上が見られなかったため、今後は、直接係り受けのある事例の精度を落とすことなく文脈情報を取り込む機構を検討したい。

謝辞 本研究は JSPS 科研費 15H01702, JP15K16045, および JST CREST (課題番号: JPMJCR1513) の支援を受けて行ったものである。また、研究の過程で有益

なコメントを頂いた東北大学の髙橋諒氏に感謝したい。

参考文献

- [1] Yuichiroh Matsubayashi and Kentaro Inui. “Distance-Free Modeling of Multi-Predicate Interactions in End-to-End Japanese Predicate-Argument Structure Analysis”. In: *COLING*. 2018, pp. 94–106.
- [2] Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. “Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis”. In: *ACL*. 2017, pp. 1591–1600.
- [3] Shuhei Kurita, Daisuke Kawahara, and Sadao Kurohashi. “Neural Adversarial Training for Semi-supervised Japanese Predicate-argument Structure Analysis”. In: *ACL*. 2018, pp. 474–484.
- [4] Tomohide Shibata and Sadao Kurohashi. “Entity-Centric Joint Modeling of Japanese Coreference Resolution and Predicate Argument Structure Analysis”. In: *ACL*. 2018, pp. 579–589.
- [5] Jörg Tiedemann and Yves Scherrer. “Neural Machine Translation with Extended Context”. In: *Proceedings of the Third Workshop on Discourse in Machine Translation*. 2017, pp. 82–92.
- [6] Longyue Wang et al. “Exploiting Cross-Sentence Context for Neural Machine Translation”. In: *EMNLP*. 2017, pp. 2826–2831.
- [7] Elena Voita et al. “Context-Aware Neural Machine Translation Learns Anaphora Resolution”. In: *ACL*. 2018, pp. 1264–1274.
- [8] 飯田 龍 et al. “述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から”. In: *自然言語処理* 17.2 (2010), pp. 25–50.
- [9] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *CVPR*. 2016, pp. 770–778.
- [10] Kyunghyun Cho et al. “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches”. In: *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. 2014, pp. 103–111.
- [11] Ashish Vaswani et al. “Attention Is All You Need”. In: *NIPS*. 2017, pp. 5998–6008.
- [12] Hirotoshi Taira, Sanae Fujita, and Masaaki Nagata. “A Japanese Predicate Argument Structure Analysis using Decision Lists”. In: *EMNLP*. 2008, pp. 523–532.
- [13] Yuichiroh Matsubayashi and Kentaro Inui. “Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis”. In: *IJCNLP*. 2017, pp. 128–133.
- [14] Diederik Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *ICLR*. 2015.