文書全体を考慮したニューラル文間ゼロ照応解析モデル

大内 啓樹　　進藤 裕之　　松本 裕治
奈良先端科学技術大学院大学 情報科学研究科
{ouchi.hiroki.nt6, shindo, matsu}@is.naist.jp

1 はじめに

日本語では述語の項が頻繁に省略される．省略された項を**ゼロ代名詞**と呼び，ゼロ代名詞が指示 (照応) している要素を**先行詞**と呼ぶ．ゼロ代名詞の先行詞を同定する処理をゼロ照応解析と呼び，特に述語と同一文外にある先行詞を同定する処理を**文間ゼロ照応解析**と呼ぶ．文間ゼロ照応解析では，先行詞同定のために文書全体を探索する必要があるため，解析が困難な問題として知られている [10, 5, 13]．

これまで文間ゼロ照応解析では，係り受け関係や大規模な格フレーム辞書などの**局所的な統語情報**が使用されてきた [8, 10, 5, 14]．しかしながら，先行詞となる単語は，述語から数文離れた位置に出現する場合が多いにも関わらず，述語-先行詞間の複数文にまたがる**大域的な文脈情報**が十分に考慮できていない．

そこで本稿では，双方向リカレントニューラルネットワーク (Bi-directional Recurrent Neural Networks; Bi-RNN)[11, 4, 3] を利用し，文書全体を考慮可能な文間ゼロ照応解析モデルを提案する．提案モデルでは，文書に含まれる各単語に割り当てられたベクトルを Bi-RNN によって伝搬させることによって，文書全体を考慮した素性ベクトルを計算する．そのベクトル表現を用いて，文書内の単語がゼロ代名詞の先行詞となる確率分布をモデル化する．

NAIST テキストコーパス [6] を用いた実験の結果，提案モデルは訓練データ内の表層的な素性情報のみの使用で，大規模外部資源を使用した既存手法と同等の解析精度を達成できることがわかった．

2 文間ゼロ照応解析

2.1 ゼロ照応の説明と分類

表 1 にゼロ照応の例を示す．ゼロ照応は (a) 文内ゼロ，(b) 文間ゼロ，(c) 外界ゼロの 3 つに大別される．文内ゼロ照応は，述語が含まれる文内にゼロ代名詞の先行詞が存在する現象である．文間ゼロ照応は，述語が含まれる文外にゼロ代名詞の先行詞が存在する現象である．外界ゼロ照応は，ゼロ代名詞の先行詞が文書内に含まれない現象である．

(a) 文内ゼロ	(ϕ ガ) 風邪を ひいた ので、 私は学校を休んだ。 [述語:ひく　ガ:私 ヲ:風邪]
(b) 文間ゼロ	彼女はパンを食べた。 (ϕ ガ) 牛乳も 飲んだ。 [述語:飲む　ガ:彼女 ヲ:牛乳]
(c) 外界ゼロ	(ϕ ガ) 映画館に 行こう かな。 [述語:行く　ガ:[一人称] ニ:映画館]

表 1: ゼロ照応の例．ϕ はゼロ代名詞を表す．

文間ゼロ照応解析では，(b) 文間ゼロと (c) 外界ゼロが解析対象となる．したがって，文内に項を持たない述語を解析対象とする．

2.2 文間ゼロ照応解析の定式化

述語 p の格 c の要素がゼロ代名詞である場合，その先行詞となる単語 y を文書 $\mathcal{X} = \{exo\} \cup \{x_t\}_1^T$ に含まれる単語から予測する．ここで，exo は先行詞が文書に含まれないこと (外界ゼロ) を表す特殊単語である．

$$\text{入力} : \mathcal{X},\ p,\ c$$

$$\text{出力} : y$$

正解の単語 y は，文書 $\mathcal{X}$ において述語 p を含む文 $p \in \mathcal{S} = (x_1, \cdots, x_n)$ 以外の単語集合 $\{\, x_t \mid x_t \notin \mathcal{S} \,\}$ に含まれる．

例えば，表 1(b) では，述語「$p =$ 飲んだ」の格「$c =$ ガ」のゼロ代名詞の先行詞を同定する．先行詞として，「彼女」を選ぶことができれば正解となる．表 1(c) では，述語「$p =$ 行こう」の格「$c =$ ガ」の先行詞は文書内にないため，正解は exo となる．

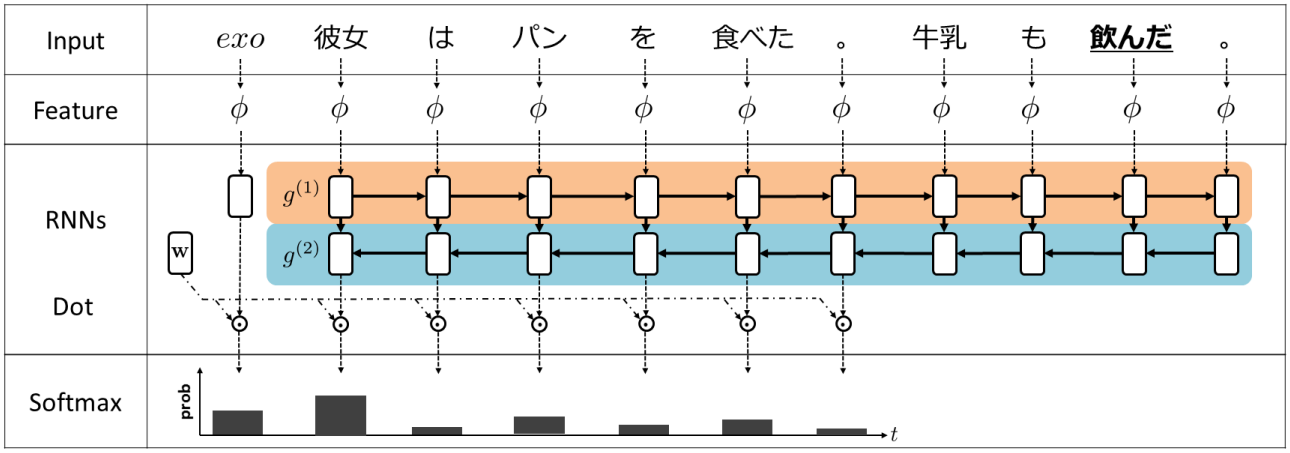

図 1: 文書エンコードモデルの構造.

3 文書エンコードモデル

提案モデルを図 1 に示す．入力として文書 $\mathcal{X} = \{exo\} \cup \{x_t\}_1^T$ と解析対象の述語 p を受け取り，文書内の単語がゼロ代名詞の先行詞となる確率分布をモデル化し，確率最大の単語を出力する．

3.1 モデルの構造

まず，文書に含まれる各単語 $x_t \in \mathcal{X}$ に対する素性ベクトル $\mathbf{x}_t$ をつくる．

$$\mathbf{x}_t = \phi(x_t, \mathcal{X}, p) \tag{1}$$

次に，双方向リカレントニューラルネットワーク (Bi-RNN) を用いて素性ベクトル $\mathbf{x}_t$ をエンコードし，新たな素性ベクトルを計算する．Bi-RNN は，L 層に積み上げられたネットワークから構成され，$l \in (1, \cdots, L)$ 層目 t 時刻の出力 $\mathbf{h}_t^{(l)}$ は以下のように計算される．

$$\mathbf{h}_t^{(l)} = \begin{cases} g^{(l)}(\mathbf{h}_t^{(l-1)}, \mathbf{h}_{t-1}^{(l)}) & (l = \text{奇数}) \\ g^{(l)}(\mathbf{h}_t^{(l-1)}, \mathbf{h}_{t+1}^{(l)}) & (l = \text{偶数}) \end{cases} \tag{2}$$

奇数層では文書を前からエンコードし，偶数層では後ろからエンコードする．どちらの層も 1 つ目の入力として $l-1$ 層 t 時刻の出力 $\mathbf{h}_t^{(l-1)}$ を受け取るが，2 つ目の入力として奇数層が l 層 $t-1$ 時刻の出力 $\mathbf{h}_{t-1}^{(l)}$ を受け取るのに対し，偶数層は l 層 $t+1$ 時刻の出力 $\mathbf{h}_{t+1}^{(l)}$ を受け取る．それら 2 つのベクトルを入力とし，遷移関数 $g^{(l)}$[1] でベクトルの更新を行う．ただし，第 1 層目のみ，関数 $g^{(1)}$ の第一引数は $\mathbf{x}_t$ をとり，$\mathbf{h}_t^{(1)} = g^{(1)}(\mathbf{x}_t, \mathbf{h}_{t-1}^{(1)})$ で計算が行われる．その後，式 2 に基づき再帰的に L 層目まで計算を行うことによって，$\mathbf{h}_t^{(L)}$ を得る．

最後に，各単語のベクトル表現 $\mathbf{h}_t^{(L)} \in \mathbb{R}^{d_h}$ を各行に持つ行列 $\mathbf{H} \in \mathbb{R}^{|\mathcal{X}| \times d_h}$ と重みパラメータベクトル $\mathbf{w} \in \mathbb{R}^{d_h}$ の積を計算し，ソフトマックス関数を用いて正規化する．

$$\mathbf{o} = \text{softmax}(\mathbf{H}\,\mathbf{w})$$

出力されたベクトル $\mathbf{o} \in \mathbb{R}^{|\mathcal{X}|}$ は，探索対象の文書 $\mathcal{X}$ に含まれる単語数 $|\mathcal{X}|$ が次元数となる．ベクトルの各要素は，各単語がゼロ代名詞の先行詞にどの程度なりやすいかを表す確率値となる．モデルパラメータは，クロスエントロピー誤差関数に基づいて学習する．

3.2 使用する素性

素性関数 ϕ により「単語/述語/距離素性」を抽出し，各単語の素性ベクトル $\mathbf{x}_t$(式 1) を計算する．「単語素性」は，各単語 x_t に割り当てられた単語 ID である．「述語素性」は，解析対象の述語 p とそのまわりの単語の単語 ID からなる素性である．「距離素性」は，単語 x_t を含む文が述語 p を含む文から何文離れているかを表す素性である．例えば，単語 x_t が 1 文前の文に含まれるなら距離 dist = 1 であり，3 文前なら距離 dist = 3 である．

それぞれの素性をベクトル表現に変換するため，埋め込みベクトル (Embeddings) を用いる．「単語/述語/距離素性」をベクトルに変換したものをそれぞれ，$\mathbf{x}_{word/pred/dist}$ と表記する．これらのベクトルを 1 本の列ベクトルに結合し，重みパラメータ行列 $\mathbf{W}_x$ との内積を計算する．

$$\mathbf{x}_t = \mathbf{W}_x\,[\mathbf{x}_{word}, \mathbf{x}_{pred}, \mathbf{x}_{dist}]$$

得られた素性ベクトル $\mathbf{x}_t$ を Bi-RNN(式 2) の入力として用いる．

[1] 本稿では Gated Recurrent Unit[2] を用いる．

	述語	文内項	文間項	外界項	項なし
学習	68,746	49,378	7,848	11,516	4
開発	13,881	10,151	1,812	1,917	1
評価	26,373	19,085	3,567	3,719	2

表 2: NAIST テキストコーパス 1.5 における述語・項 (ガ格) の分布.「文内項」は直接係り受け関係にある項と文内ゼロ照応の項を合わせた数で,「文間項」と「外界項」はそれぞれ文間・外界ゼロ照応の関係にある項の数である.

4 実験

日本語では，他の格に比べ，ガ格の項がゼロ代名詞である頻度が特に多い [7] ため，本稿ではガ格の文間ゼロ照応解析に着目し実験を行う.

4.1 実験設定

文間ゼロ照応解析研究は以下の 2 つの異なる実験設定で行われてきた:

1. 文間ゼロ照応の単独解析
2. 文内述語項構造と文間ゼロ照応の同時解析

2 つの実験設定の主な違いは，**文内項判定**を行うか否かである．文内項[2]判定とは，解析対象の述語の項が述語と同一文内に存在するか否かを判定することである.

文間ゼロ照応を単独で解析する場合，文内述語項の正解アノテーションに基づいて，文内項を持たない述語のみを解析対象とする．つまり，文内項判定として正解データを用いる．一方，文内述語項構造との同時解析では，すべての述語を解析対象とし，文内述語項と文間ゼロ照応を同時に解く．したがって，文内項判定もモデルの予測に基づいており，文間ゼロ照応の単独解析より困難な問題であると言える.

本稿では，両方の設定で実験を行い，モデルの解析性能を報告する．提案モデルは文間ゼロ照応の単独解析だけでなく，文内述語項との同時解析にも適用することが可能である．具体的には，まず，述語を含む文とそれ以外のすべての文に含まれる単語の確率分布を計算する．その結果，述語と同一文内の単語が確率最大になった場合，文間ゼロ照応の先行詞は *null* として出力する.

4.2 実験データ

NAIST テキストコーパス 1.5 [6] を用いる．実験では，次に示すような，標準的なデータ分割法 [12] を採用し，モデルの訓練・開発・評価を行った.

訓練: 1 月 1-11 日の記事と，1 月から 8 月の社説.

開発: 1 月 12，13 日の記事と，9 月の社説.

評価: 1 月 14-17 日の記事と，10 月から 12 月の社説.

表 2 に，訓練・開発・評価データにおけるガ格の項の分布を示す．[3]

実験で用いる単語境界として，NAIST テキストコーパスの正解アノテーションを利用する．モデルの学習に，外部資源は一切利用しない.

4.3 実装詳細

提案モデルの実装は，深層学習ライブラリ Theano [1] を利用し，CPU(Intel 6 Core Xeon E5-4617) 上で実行する．使用する単語ベクトルとして，訓練データ内に出現する単語の中で，頻度 2 以上のもののみを使用し，残りの単語は未知語として未知語のベクトルに割り当てる.

エポック数は 50 に設定し，開発データの F 値が最も良いエポックでの評価データの結果を報告する．パラメータの最適化は確率的勾配降下法 (SGD) で行う．学習係数は Adam [9] を用いて自動調整する.

各ハイパーパラメータは，以下のように選ぶ.

埋め込みベクトル : 32 次元に設定し，初期値として $[-0.01, 0.01]$ から一様分布に従ってサンプリングした値を設定する.

重みパラメータ行列 : 32×32 とし，$[-0.01, 0.01]$ から一様分布に従ってサンプリングした値を設定する.

正則化項 : 正則化項のハイパーパラメータ λ は $[0.001, 0.0005, 0.0001]$ の中から，開発データの F 値が最大となるものを選ぶ.

4.4 比較手法

NAIST テキストコーパス 1.4β を用いて文間ゼロ照応解析を行っている先行研究と比較を行う．本稿で用いるコーパスとバージョンが異なるため，厳密な比較はできないが参考のため比較する．また，先行研究では，格フレームなどの外部資源や品詞・構文情報を素性として使用しているため，本稿の設定より有利な設定であると言える.

[2]直接係り受けのある項と文内ゼロ照応の項の両方を指す.

[3]「項なし」にカウントしているものは，アノテーションミスであることがわかった.

	開発	評価
提案モデル (単独)	31.7	27.6
提案モデル (同時)	23.7	21.7
Taira et al. (2008) [12]	-	23.5
Imamura et al. (2009) [8]	-	13.1
Sasano & Kurohashi (2011) [10]	-	**24.4**
林部ら (2014) [14]	-	19.6

表 3: 文間ゼロ照応解析 (ガ格) の F 値．提案モデル (単独) は，文間ゼロ照応解析のみを行った結果であり，提案モデル (同時) は文内項と文間ゼロ照応を同時に解析を行った結果である．比較している既存研究では，訓練データの他に外部資源を用いている．

4.5 結果

表 3 に，解析結果を載せる．文間ゼロ照応を単独で行った場合の提案モデル (文内項判定:正解) は，評価データにおいて 30%にせまる F 値を記録している．このモデルは，文内項を持たない述語のみを解析対象としているため，先行研究のモデルと厳密に比較することはできないが，それらと比べ高い解析性能を表している．この解析精度は文内項判定にすべて正解した場合の文間ゼロ照応解析精度を表しており，この精度をさらに引き上げることが今後の課題の 1 つとなる．

また，文内項判定と文間ゼロ照応解析を同時に行った結果も報告する．提案モデルは，外部資源や構文情報を用いていないにも関わらず，先行研究と同等の精度を達成している．この結果から，Bi-RNN で文書の大域的な文脈を考慮することによって，単語などの表層情報のみからでも効果的な学習が行えることがわかる．今後，外部資源などを用いて，さらにモデルの性能が改善可能かどうかを調査をしていきたい．

5 おわりに

本稿では，文間ゼロ照応解析に取り組んだ．双方向リカレントニューラルネットワークを用いて，文書全体を考慮可能な文間ゼロ照応解析モデルを提案した．NAIST テキストコーパスを用いた実験を行い，提案モデルは外部資源を使用せずに既存手法と同等の精度を達成した．このことから，文書内の大域的な文脈情報をとらえることの重要性を確認した．今後の課題として，(i) 外部資源を用いたより効果的なモデルの学習法の開発と，(ii) 効率的な文内項・文間ゼロ照応の同時解析モデルの開発が挙げられる．

参考文献

[1] Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian J. Goodfellow, Arnaud Bergeron, Nicolas Bouchard, and Yoshua Bengio. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.

[2] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of EMNLP*, pages 1724–1734, 2014.

[3] Alan Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. Hybrid speech recognition with deep bidirectional lstm. In *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop*, 2013.

[4] Alex Graves, Santiago Fernández, and Jürgen Schmidhuber. Bidirectional lstm networks for improved phoneme classification and recognition. In *International Conference on Artificial Neural Networks*, pages 799–804, 2005.

[5] Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. Japanese zero reference resolution considering exophora and author/reader mentions. In *Proceedings of EMNLP*, pages 924–934, 2013.

[6] Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. Annotating a japanese text corpus with predicate-argument and coreference relations. In *Proceedings of the Linguistic Annotation Workshop*, pages 132–139, 2007.

[7] Ryu Iida, Kentaro Torisawa, Jong-Hoon Oh, Canasai Kruengkrai, and Julien Kloetzer. Intra-sentential subject zero anaphora resolution using multi-column convolutional neural network. In *Proceedings of EMNLP*, pages 1244–1254, 2016.

[8] Kenji Imamura, Kuniko Saito, and Tomoko Izumi. Discriminative approach to predicate-argument structure analysis with zero-anaphora resolution. In *Proceedings of ACL-IJCNLP*, pages 85–88, 2009.

[9] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980*, 2014.

[10] Ryohei Sasano and Sadao Kurohashi. A discriminative approach to japanese zero anaphora resolution with large-scale lexicalized case frames. In *Proceedings of IJCNLP*, pages 758–766, 2011.

[11] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, pages 2673–2681, 1997.

[12] Hirotoshi Taira, Sanae Fujita, and Masaaki Nagata. A japanese predicate argument structure analysis using decision lists. In *Proceedings of EMNLP*, pages 523–532, 2008.

[13] 松林優一郎，中山周，乾健太郎．日本語述語項構造解析タスクにおける項の省略を伴う事例の分析．**自然言語処理**, pages 433–463, 2015.

[14] 林部祐太，小町守，松本裕治．述語と項の位置関係ごとの候補比較による日本語述語項構造解析．**自然言語処理**, pages 3–25, 2014.