Vol.2018-NL-238 No.5
2018/12/11

大規模格フレームによる解候補削減を用いたニューラルネットゼロ照応解析

山城 颯太[1,a)]　西川 仁[1,b)]　徳永 健伸[1,c)]

概要：本論文では日本語文内・文間ゼロ照応解析モデルを提案する．文間ゼロ照応解析において複数格の同時推定を行う際，複数の文をまたぐ大量の格要素の組合せ候補を取り扱う必要があり，これはゼロ照応解析モデルの訓練，解析に際して重大な障害となる．この問題に対して，我々は格フレームの情報を用いた効果的な解候補削減手法を提案する．また文間ゼロ照応解析に対して，モデルが解析対象動詞から離れた文脈も考慮できるよう，ローカルアテンション付き RNN を導入した．日本語均衡コーパスを用いて提案モデルを評価し，解候補削減を用いることで 0.056 の精度向上を確認した．また，ローカルアテンション付き RNN を導入することで，文間ゼロ照応解析の精度が上昇することも確認した．

1. はじめに

ゼロ照応解析とは，テキスト中の述語の省略された項 (ゼロ代名詞) を検出し，項として埋めるべき格要素を同定するタスクである．格要素は先行詞としてテキスト中で言及されている場合もあれば，言及されていない場合もある．前者の場合，先行詞は述語と同じ文中にある（文内ゼロ照応）か，先行する文中にある（文間ゼロ照応）[*1]．後者 (外界ゼロ照応) の例として，テキストの著者である主語が明示的に言及されない場合などがある．

(1) 大岡山商店街でも (ϕ ガ) お洒落な建物を見かけるようになった．カフェテリアが特に多くて，今月も新しく (ϕ ガ)(ϕ ニ)オープンしてる．

例 (1) では「見かける」のガ格と「オープンしてる」のガ格，ニ格が省略されている．「オープンしてる」のガ格の格要素は同文中に言及されている 「カフェテリア」 であり (文内ゼロ照応)，ニ格の格要素は前文で言及されている「大岡山商店街」である (文間ゼロ照応)．一方，「見かける」 のガ格の格要素はテキスト中では明示的に言及されていない著者である (外界ゼロ照応)．

本論文では特に日本語のゼロ照応解析を取り扱うが，項の省略が起こる pro-drop 言語は日本語だけではなく，他に中国語，イタリア語，スペイン語などがあり，各語で日本語ゼロ照応解析と類似したタスクに取り組む研究が数多くある (Yin et al., 2017; Chen and Ng, 2016; Iida and Poesio, 2011; Rello et al., 2012)．また英語では意味役割付与タスクがゼロ照応解析に似た研究として挙げられる (Zhou and Xu, 2015; He et al., 2017)．日本語ゼロ照応解析は，日本語述語項構造解析の部分問題であり，自動要約 (Yamada et al., 2017) や情報抽出 (Sudo et al., 2001)，機械翻訳 (Kudo et al., 2014) など様々な自然言語処理アプリケーションの精度改善にとって重要であるため，緊急に解決されるべき課題として盛んに研究されている (Sasano and Kurohashi, 2011; Hangyo et al., 2013; Ouchi et al., 2017; Hangyo et al., 2013; Matsubayashi and Inui, 2017)．

本研究の貢献は大きく二つに分けられる．第一に大規模均衡コーパス上で日本語ゼロ照応解析を行い評価したことと，第二にこの大規模均衡コーパス上で文内・文間ゼロ照応解析を可能にするための解候補削減手法を提案したことの二点である，

従来のゼロ照応解析研究は，新聞記事のみからなる『NAIST テキストコーパス』(NTC) (Iida et al., 2007) で評価を行うものが多かった．従って，それらの評価ではテキストドメインの違いによる影響が考慮されていない．しかしゼロ照応解析結果の応用を考えた時，新聞のみならずブログ，QA，書籍，白書，雑誌などあらゆるドメインの文書に対して頑健なゼロ照応解析手法こそより有用性が高い．我々は『 現代日本語書き言葉均衡コーパス』(BCCWJ) (Maekawa et al., 2014) を評価実験に使用した．BCCWJ は 13 ドメインにまたがって構築された約一億語からなる日本語均衡コーパスである．このうちの約 100 分の 2 にあたる約二百万語からなるコアデータに対しては，

[1] 東京工業大学 情報理工学院
[a)] yamashiro.s.aa@m.titech.ac.jp
[b)] hitoshi@c.titech.ac.jp
[c)] take@c.titech.ac.jp
[*1] この研究では後方照応は扱わない

ⓒ 2018 Information Processing Society of Japan

距離	ガ格	ヲ格	ニ格	total	%
0	16,621	4,545	2,059	23,225	50.4
1	8,231	1,764	1,113	11,108	24.1
2	3,396	599	430	4,425	9.6
3	1,792	317	227	2,336	5.1
4	1,020	172	126	1,318	2.9
5	690	83	84	857	1.9
6	414	45	51	510	1.1
≥ 7	1,917	217	178	2,312	5.0
total	34,081	7,742	4,268	46,091	

表 1 格要素と述語の距離の分布

距離	OW	PB	PN	PM	OC	OY
0	72.3	49.5	51.1	40.3	38.8	49.8
1	15.1	25.0	24.4	23.9	29.1	23.1
2	5.6	9.8	9.6	11.2	13.4	8.7
3	2.4	5.0	4.8	6.7	6.9	4.5
4	0.9	2.7	2.6	4.3	3.9	3.5
5	0.9	1.3	2.3	2.7	2.9	1.6
6	0.4	1.1	1.0	1.7	1.4	1.7
≥ 7	2.5	5.5	4.4	9.3	3.6	7.2

OW: 白書, PB: 書籍, PN: 新聞,
PM: 雑誌, OC: QA, OY: ブログ

表 2 文書ドメインごとのガ格ゼロ照応の分布 (%)

人手による述語項構造と照応関係の付与がされている．また，BCCWJ は新聞，雑誌，書籍，白書，Yahoo!知恵袋，Yahoo!ブログの 6 ドメインにまたがったテキストを含んでいる．ドメインによるゼロ照応解析の性能の違いを調べるために，我々は BCCWJ を使用した．

表 1 は BCCWJ コアデータセットの述語と格ごとの格要素の距離の分布を示している．ここでの距離は述語と格要素の間の文数である．距離 0 は文内照応を示しており，距離 1 以上は文間ゼロ照応を示している．この表から，半数以上のゼロ照応が文間ゼロ照応であることがわかる．表 2 はテキストドメインごとに分類した述語とガ格の格要素との距離の分布を示している．この表から，文内，文間ゼロ照応のドメインごとの違いが確認できる．これらの観察から，異なるタイプのテキスト上で評価実験を行うことの重要性が示唆される．

表 1 に示すとおり現実の文書には文間ゼロ照応が頻出するが，従来のゼロ照応解析研究の多くは，文内ゼロ照応のみに焦点を絞っている (Iida et al., 2015; Shibata et al., 2016; Ouchi et al., 2017; Matsubayashi and Inui, 2017)． Ouchi et al. (2017) は，文内ゼロ照応のみを取り扱う理由として，探索範囲の問題を指摘している．文間ゼロ照応では，格要素候補をテキスト全体から探す必要があるため，文内ゼロ照応解析に比べて探索範囲が拡大する． Matsubayashi and Inui (2017) は解析に際して文脈素性を取り入れるために，リカレントニューラルネットワーク (RNN) を導入し，格要素候補と述語が含まれる文を読み込ませている．しかしこれと同じ手法を文間ゼロ照応解析において適用しようとすると，テキスト全体を RNN に入力として与える必要がある．長距離の文脈を記憶する仕組みを持つ LSTM や GRU を使用しても，システムがテキスト全体における長距離の依存関係を十分に学習できるとは限らない．また，テキスト全体を記憶しなくても，選択的に抽出された文脈情報のみで解析できる可能性がある．

先述の研究と異なり， Sasano and Kurohashi (2011) と Hangyo et al. (2013) は，文内・文間ゼロ照応解析手法を提案している．しかし彼らはそれぞれ独自に収集，アノテーションを行った Web コーパスを用いて評価実験を行っている．

これら 2 つの問題に対して，本研究では様々なドメインの文書への対応を可能とするために大規模格フレームを利用し，述語が取りうる複数の格要素の組合せから最適なものを選ぶ．解候補となる格要素の組合せが膨大になる問題に対しては，格フレームを使用した解候補削減手法を取り入れることで，より汎用性の高い文内・文間ゼロ照応解析モデルを提案する．

ひとつのモデルで文内・文間のゼロ照応解析を同時に行う際，各格に対してそれぞれ独立に解析を行うより，他の格の情報を利用して複数格を同時に解析する方がより良い精度が得られると考えられる．しかし複数格を同時に解析する際には，先行詞の広大な探索範囲の問題に対処する必要がある．特に機械学習を適用する際，正解の候補となる名詞の組合せが大幅に増加することから，BCCWJ の場合では正例と負例の比率が約 1 対 20,000 と著しく不均衡となる．このような偏った訓練データは不必要に計算量を増幅させ，かつモデルの汎化を妨げる要因となる．我々は，学習に不要な負例を削減するために，解析対象述語に対応する格フレームを用いた効率的な解候補削減手法を提案する．この提案手法により，正解を候補に残しつつ，約 1,000 分の 1 にまで解候補を削減することに成功した．また，我々は RNN にローカルアテンション機構 (Luong et al., 2015) を導入することで，前文中のどの部分に注意を向けて解析するかをシステムに学習させた．なお，BCCWJ を用いた文内・文間のガヲニ格を対象とするゼロ照応解析は．本研究が初の試みである．

2. 関連研究

2.1 日本語ゼロ照応解析

表 3 は，タスクの種類，使用しているコーパスのドメイン，コーパスのサイズ，手法の観点から関連研究をまとめたものである．Hangyo et al. (2013) はランキング SVM を用いて，Web コーパスに対して文内，文間，外界のゼロ照応解析を同時に行っている．この Web コーパス

	タスク				ドメイン			サイズ	手法			解候補
	係り受け	文内	文間	外界	新聞	Web	etc.	(文数)	線形	NN	+att	
(Imamura et al., 2009)	o	o	o		o			40,000	o			独立
(Hangyo et al., 2013)		o	o	o		o		3,000	o			組合せ
(Ouchi et al., 2015)	o	o			o			40,000	o			組合せ
(Shibata et al., 2016)	o	o				o		15,000		o		組合せ
(Iida et al., 2016)		o			o			40,000		o		独立
(Ouchi et al., 2017)	o	o			o			40,000		o		独立
(Matsubayashi and Inui, 2017)	o	o			o			40,000		o		独立
(Sasano and Kurohashi, 2011)		o	o			o		1,000	o			組合せ
提案手法		o	o		o	o	o	60,000	o	o	o	組合せ

表 3 関連研究

は 1,000 文書からなり，それぞれ Web ページの冒頭 3 文を抜き出したものである (Hangyo et al., 2012)． Shibata et al. (2016) はフィードフォワードニューラルネットワーク (FNN) を用いて，Web コーパス (Hangyo et al., 2012) に対して直接の係り受け関係と文内のゼロ照応解析を同時に行っている． Matsubayashi and Inui (2017) はフィードフォワードニューラルネットワーク (FNN) とリカレントニューラルネットワーク（RNN）を組合せて用いることで，NTC に対して直接の係り受け関係と文内のゼロ照応解析を同時に行い，直接の係り受け関係と文内ゼロ照応解析の state-of-the-art を達成した． Sasano and Kurohashi (2011) は対数線形モデルを用いて，979 文からなる Web コーパスに対して文内と文間のゼロ照応解析を同時に行い，文内・文間ゼロ照応解析の state-of-the-art を達成した．これらに対して我々は，ランキング SVM[*2] (Joachims, 2006) モデルと FNN と RNN の組合せモデルを用いて，『現代日本語書き言葉均衡コーパス』 (BCCWJ) (Maekawa et al., 2014) に対して，文内・文間のゼロ照応解析を同時に行う．

2.2 大規模格フレーム

格フレームとは述語とその述語が取りうる項を述語の格パターンごと，格ごとに整理した共起情報である．表 4 のように格パターンに基づいて格フレームを分けることで，述語と項間の語彙的選好の知識を照応解析に利用することができる (Sasano et al., 2008; Sasano and Kurohashi, 2011; Hangyo et al., 2013)．格フレームの構築に関しては Kawahara and Kurohashi (2006) が Web テキストから格フレームを自動構築する手法を提案している．これらの大規模 Web コーパスから取得，整理された格フレーム知識は京大格フレーム[*3]として公開されている．

[*2] https://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

[*3] http://www.gsk.or.jp/catalog/gsk2008-b/ ただし，リンク先の京大格フレームは古い版であり，本項において使用したものは未公開の新しい版である．

2.3 解候補削減

文間ゼロ照応に際して，いくつかの先行研究ではそれぞれに解候補削減の基準を設定している． Sasano and Kurohashi (2011) と Hangyo et al. (2013) は述語の複数の格を同時に推定しており，述語が含まれる文より 3 文前までに出現する格要素の先行詞候補をすべて含めている．3 文より前の文にも先行詞は出現しうるが，Hangyo et al. (2013) は NTC の述語に対する格要素のうち 82.9%が 3 文中に出現すると報告した．Imamura et al. (2009) は述語の複数の格をそれぞれ独立に推定しており，一文前までに現れる述語の格要素として選ばれた名詞のみを解析の対象としている．この制限によって，NTC 中，何も制限のない状態では平均 102.2 語の名詞を解候補としなければならなかったところを平均 3.2 語まで抑え，ゼロ代名詞の格要素のうち 62.5%をカバー出来たと彼らは報告している．Ouchi et al. (2015) は述語項構造解析を，複数述語とその項候補の二部グラフとして定式化し，その局所解を山登り法で探索している．

3. 提案モデル

本研究の提案手法は二つの構成要素からなり，一つは格フレーム内の単語分散表現を使用した解候補削減アルゴリズムで，もう一つは解候補削減に使用した分散表現を利用するニューラルネットゼロ照応解析モデルである[*4]．

3.1 モデル

解析対象述語 p が含まれる文を S_0 とし， 入力文書 t に含まれる S_0 から h 文前までの文をそれぞれ $S_{-1}, S_{-2}, \cdots, S_{-h}$ とする．S_0 から S_{-h} までに含まれるすべての名詞の集合を $E_p = \{e_1, e_2, \cdots, e_n\}$ とする．これらに加えて『照応なし』または『外界照応』を意味する e_{none} を E_p に追加する．述語 p に対応する京大格フレーム中の格フレーム群を $CF_p = \{cf_1^p, cf_2^p, \cdots, cf_m^p\}$ とする．1 つの格フレーム cf_l^p には，それぞれの格 $c \in$ { ガ格, ヲ格, ニ格 }

[*4] https://github.com/yamashiros/Japanese_zero_anaphora

格フレーム	ガ格	出現数	ヲ格	出現数	ニ格	出現数
オープンしてる:動 1	店	129	–	–	近く	6
	カフェ	38	–	–	跡地	2
	レストラン	14	–	–	ところ	2
	...	...	–	–	...	...
オープンしてる:動 2	ブランド	12	ショップ	59	–	–
	専門家	8	サロン	18	–	–
	オーナー	4	ブティック	13	–	–
	...	...	...	...	–	–

表 4 「オープンしてる」の格フレーム例

に対応する 3 つの格スロットがあり，E_p 中に含まれるいずれかの名詞がそれぞれの格スロットに対応する格要素である．格スロットと格要素の対応付けを $a = \langle$ ガ格 $\leftarrow e_i$, ヲ格 $\leftarrow e_j$, ニ格 $\leftarrow e_k \rangle$ とする．述語項構造候補を (cf_l^p, a) とし，これを表現する素性ベクトルを $\boldsymbol{f}(cf_l^p, a, t)$ とする．このモデルの出力は以下の式 (1) で表せる．$\boldsymbol{w}$ は訓練データから学習されるパラメータである．このモデルは，Hangyo et al. (2013) のモデルをベースとしている．

$$cf_l^{p*}, a^* = \underset{cf_l^p, a}{\operatorname{argmax}} \boldsymbol{w} \cdot \boldsymbol{f}(cf_l^p, a, t) \tag{1}$$

3.2 素性

素性ベクトル $\boldsymbol{f}(cf_l^p, a, t)$ は以下 5 タイプの素性の組合せからなる: **ベースモデル素性，格要素分散表現，述語分散表現，格フレーム内平均ベクトル** (MVC)，**文脈ベクトル**．

3.2.1 ベースモデル素性

ベースモデル素性 ϕ_{BMF} の各要素は実数かバイナリ値である．ベースモデル素性 ϕ_{BMF} は Sasano et al. (2008) の確率的格解析モデルから得られる表層の係り受けの確率と Hangyo et al. (2013) が提案する素性群からなる．Hangyo et al. (2013) の素性は格フレーム素性，述語素性，文脈素性の 3 種類からなる．例えば，ある格要素がその格フレームの格スロットに埋まるかどうかの確率は格フレーム素性の一つである．

3.2.2 格要素分散表現

格要素分散表現 ϕ_e は各格 c の格要素 e_c に対応する 3 つの分散表現から構成される．語の分散表現を生成するモデルとしては word2vec (Mikolov et al., 2013) を使用した[*5]．

3.2.3 述語分散表現

述語分散表現は word2vec を使って生成された解析対象述語の単語分散表現である．

3.2.4 格フレーム内平均ベクトル（MVC）

表 4 に示すように京大格フレーム内では，述語 p に対するそれぞれの格フレーム cf_l^p は各格 c に対応する単語リストから構成される．例えば，「オープンしてる:動 1」のガ格には『店』，『カフェ』，『レストラン』などが格納されている．$W_{cf_l^p(c)}$ を格フレーム cf_l^p と格 c に対応して京大格フレーム中に出現する格要素の全体とする．例えば $W_{\text{オープンしてる:動}_1\text{(ガ)}}$ の要素は前述の『店』，『カフェ』，『レストラン』などである．

ϕ_w を語 $w \in W_{cf_l^p(c)}$ の分散表現ベクトル，$\mathrm{count}(cf_l^p, c, w)$ を語 w が格フレーム cf_l^p の格 c の格要素として出現する回数とする．この時，格フレーム内平均ベクトル (MVC)$\overline{\phi}_{cf_l^p(c)}$ は，格フレーム cf_l^p 中の各格 c の分散表現ベクトル $W_{cf_l^p(c)}$ の重み付き平均として計算される．

$$\overline{\phi}_{cf_l^p(c)} = \frac{\sum_{w \in W_{cf_l^p(c)}} \mathrm{count}(cf_l^p, c, w) \cdot \phi_w}{\sum_{w \in W_{cf_l^p(c)}} \mathrm{count}(cf_l^p, c, w)} \tag{2}$$

例えば図 4 において，格フレーム「オープンしてる:動 1」はガ格に『店』を 129 回取っているので，$\overline{\phi}_{\text{オープンしてる:動}_1\text{(ガ)}}$ は以下のように計算される．

$$\overline{\phi}_{\text{オープンしてる:動}_1\text{(ガ)}} = \frac{129 \cdot \phi_{\text{店}} + 38 \cdot \phi_{\text{カフェ}} + \cdots}{129 + 38 + \cdots} \tag{3}$$

$\overline{\phi}_{cf_l^p(\text{ガ})}$，$\overline{\phi}_{cf_l^p(\text{ヲ})}$，$\overline{\phi}_{cf_l^p(\text{ニ})}$ を結合して $\overline{\phi}_{cf_l^p}$ を生成する．$\overline{\phi}_{cf_l^p}$ を使って，a と cf_l^p の関連（選択選好）を測り，尤もらしい組合せを探索する．なお我々は，MVC を照応解析，解候補削減の両方で使用する．

3.2.5 文脈ベクトル

文脈ベクトル $c_{cf_l^p,a,t}$ はローカルシングルアテンション機構付き RNN の出力である．この RNN は解析対象述語を含んだ文とその前方 h 文を受け取り，対象述語に対する文脈をモデリングする．$\mathrm{Enc}(S_{-h:0})$ を，$S_{-h:0}$ を入力として与えられた時の RNN エンコーダの隠れ状態とする．$\mathrm{LocalAtt}(\cdot)$ はローカルシングルアテンション機構を表す．

我々のアテンション機構モデルは他の素性ベクトルの連結に基づいてアライメント重みベクトルを推論する．文脈ベクトル $c_{cf_l^p,a,t}$ はこのアライメント重みベクトルによる，エンコーダの出力 $\mathrm{Enc}(S_{-h:0})$ の重み平均として計算される．

[*5] 日本語 wikipedia (2016-09-20) の本文全文から取得した約 100 万記事に対して，次元数を 500，window を 15 として学習させることで得られたモデルを使用した．

$$c_{cf_l^p,a,t} = \mathrm{LocalAtt}([\phi_{BMF}; \phi_e; \overline{\phi}_{cf_l^p}], \mathrm{Enc}(S_{-h:0})) \quad (4)$$

直感的には，この機構は我々のモデルがアライメントベクトルを介して述語から離れた長文脈中の語を格要素として識別することを可能にしている．格フレームがその動詞から離れた名詞を格要素として取るようなケースに対して，我々はこのメカニズムが直接的にその現象をモデル化することを期待している．

4. 格フレーム中の分散表現を利用した解候補削減

ゼロ代名詞となる格要素の先行詞候補を網羅的に探索すれば，列挙される述語項構造候補 (cf_l^p, a) の集合は爆発的な規模となり，探索範囲は非実用的なものとなる．Sasano and Kurohashi (2011) の基準を参考に，ゼロ代名詞となる格要素の先行詞候補は述語が含まれる文より 3 文前までのみを範囲として解候補削減を行っている．つまり 3.1 の h を 3 とした．BCCWJ 中の格要素の分布は表 1 のようになっているため，この制限によってゼロ代名詞の 89.16%をカバーできることがわかる．

n と m をそれぞれ E_p 中の名詞句数，対象述語の格フレーム数とすると，この制限を用いてもなお，候補の数は $O(n^3m)$ となり，BCCWJ 中の各動詞に対して約 20,000 個の述語項構造候補が出現する．

4.1 述語内平均ベクトル（MVP）

我々は，格フレーム候補と項候補の組合せについて 3.2.4 で提案した MVC と，**述語内平均ベクトル** (MVP) $\overline{\phi}_{p(c)}$ の二種類の平均ベクトルを使用した効率的な解候補削減手法を提案する．MVP は各格 c について述語 p に対応するすべて格フレームに渡って MVC $\overline{\phi}_{cf_l^p(c)}$ の重み平均を取ったベクトルである．重みは京大格フレーム中の各格フレームの頻度に基づく．我々の解候補削減手法は Ouchi et al. (2015) の山登り法を参考に，格フレーム候補と項候補の組合せ数を削減する．この解候補削減は計算効率のみを目的とするのではなく，訓練データ中の正例・負例のデータ数の非対称性の解消も目的とする．我々のケースでは，1 つの正例に対して 20,000 の負例が生じるため，これに対処している．前述したように，我々は訓練データ中のほとんどの負例は訓練に貢献しないと考え，解候補削減を行う．

4.2 アルゴリズム

我々の提案する解候補削減手法をアルゴリズム 1 に示す．ある述語 p には，文脈に対するその語義の曖昧性を反映した複数の格フレーム CF_p が存在する．それぞれの格フレーム cf_l^p に対応する格フレーム内平均ベクトル $\overline{\phi}_{cf_l^p(c)}$ はその格フレームの選択選好を反映しているため，これと項候補ベクトル ϕ_e の距離が近いほど，その項候補 e は対象格フレーム cf_l^p の格スロット c に埋まりやすいと言える．このアルゴリズムは与えられた述語に対して，二つのベクトル間の距離が最も近くなる格フレームと項候補の組合せを探索する．しかしながら，京大格フレームは自動的手法で構築されているので，本来別々の格フレームが一つの格フレームとしてまとめられてしまっている，あるいは同じ一つの格フレームが別々に分断されてしまっている可能性がある．この問題に対処するために，我々は提案する解候補削減手法に二種類の平均ベクトルを導入した．MVC はある述語に対する格フレームの違いを区別し，MVP は格フレームの違いを考慮せず述語のみを考慮する．

アルゴリズム 1 解候補削減アルゴリズム

Input:
a predicate p to be analyzed,
a set of case frames CF_p corresponding to p,
a set of cases $C = \{$ ガ格, ヲ格 , ニ格 $\}$,
a set of nouns E_p appearing within the h preceding sentences.
Output:
optimal cf_l^{p*}, e_c^* for the analyzed p and each case $c \in C$.

1: **for** each case $c \in C$ **do**
2: $e_c^{(0)} \Leftarrow \underset{e \in E_p}{\mathrm{argmax}} \cos(\bar{\phi}_{p(c)}, \phi_e)$ ▷ $\bar{\phi}_{p(c)}$ is the MVP
3: **end for**
4:
5: $cf^{(0)} \Leftarrow \underset{cf_l^p \in CF_p}{\mathrm{argmax}} \sum_{c \in C}$ Pseudo-Score$(cf_l^p, e_c^{(0)})$
6: $t \Leftarrow 0$
7: **repeat**
8: **for** each case $c \in C$ **do**
9: $e_c^{(t+1)} \Leftarrow \underset{e_c \in E_p}{\mathrm{argmax}}$ Pseudo-Score$(cf^{(t)}, e_c)$
10: **end for**
11:
12: $cf^{(t+1)} \Leftarrow \underset{cf_l^p \in CF_p}{\mathrm{argmax}} \sum_{c \in C}$ Pseudo-Score$(cf_l^p, e_c^{(t+1)})$
13: $t \Leftarrow t + 1$
14: **until** $e_c^{(t)} = e_c^{(t+1)}$ and $cf^{(t)} = cf^{(t+1)}$
15: **return** $cf_l^{p*} \Leftarrow cf^{(t)}, e_c^* \Leftarrow e_c^{(t)}$ for each case $c \in C$
16:
17: **function** Pseudo-Score(cf_l^p, e)
18: $score \Leftarrow 0$
19: **for** each case $c \in C$ **do**
20: $score \Leftarrow score + P(p, cf_l^p, e, c)$
21: $score \Leftarrow score + \cos(\bar{\phi}_{cf_l^p(c)}, \phi_e)$
22: $score \Leftarrow score + 0.5 \times d_{p,e}$ ▷ $\bar{\phi}_{cf_l^p(c)}$ is the MVC
23: **end for** ▷ $d_{p,e}$ is the distance between p and e
24:
25: **return** $score$
26: **end function**

まず初期値として各格 $c \in C$ に埋まりうる項 $e_c^{(0)}$ を仮に定める (行 1-3)．MVP $\overline{\phi}_{p(c)}$ と項候補の分散表現 ϕ_e とのコサイン距離を求め，これが最小となる，すなわち対象述語に埋まる項群に最も近い項を初期項とする．この段階では，MVP を使用することで特定の格フレームではなく述語のみを考慮している．格フレーム候補と初期項の組合せを入力とした Pseudo-Score(行 17-26) の返すスコアに基づいて，これらの初期項に対して最適な格フレーム $cf^{(0)}$ を格フレームの初期値とする (行 5)．Pseudo-Score については Sasano et al. (2008) を参考に，我々は以下の 3 つの要素を考慮した．(1) 京大格フレームに基づく (述語，格フレーム，深層格，項) の組合せの出現確率，(2) 格フレーム

内平均ベクトル（MVC）と項候補の間のコサイン類似度，および (3) 述語と項候補の間の文数，である．このスコアの係数は経験的に定めた．以降，格フレーム $cf^{(t)}$ を固定して項 $e_c^{(t+1)}$ を探索するフェーズ (行 8-10) と項 $e_c^{(t+1)}$ を固定して格フレーム $cf^{(t+1)}$ を探索するフェーズ (行 12) を繰り返し，格フレームと項が更新されなくなればループを抜ける (行 6-15)．このアルゴリズムでは返り値として最もスコアの高い格フレームと項の組合せを返すが，実際にはループ中の毎回の探索過程で計算した項候補のうち 3 ベストまでを候補として保存する．最終的な出力は探索の過程で保存されたすべての格フレームと項の組合せである．提案した解候補削減手法により，約 70%の正解を候補に残しつつ，約 1,000 分の 1 まで解候補を削減することができた．

5. 評価実験

5.1 ゼロ照応解析手法

学習手法にはランキング SVM と FNN を使用しそれぞれ比較した．先行研究 (Sasano and Kurohashi, 2011; Hangyo et al., 2013) と同様に，まず文書全体に対して形態素解析，固有表現抽出，構文解析を行う．これには JUMAN Ver.7.01[*6], KNP Ver.4.16[*7], CaboCha Ver.0.69[*8] を用いた．

5.1.1 S0

提案する解候補削減を行い，ベースモデル素性を使用して SVM モデルを実装した．ランキング学習には SVM^{rank} (Joachims, 2006) を使用した．カーネルは線形である．このモデルは正例と負例から識別関数を学習し，この識別関数が最も高い解候補を一つ出力する．

5.1.2 S0_each

提案する解候補削減手法の効果を評価するためには，解候補削減を用いないモデルと比較することが自然である．しかしながら，前方 3 文までに先行詞候補の探索範囲を制限しても，述語一つあたりに対して 20,000 の述語項構造候補が出現するため，訓練時の計算複雑性は現実的ではない．これは複数の格を同時推定するために，格要素候補同士の組合せを考慮していることが原因である．そこで我々は，それぞれの格を独立に解析することで，解候補削減が必要ない単一格解析手法を用意した．この手法では，3 つの格に対してそれぞれ別の SVM モデルを用意し，これらを独立に学習させて，評価の際は各格に対応するモデルのそれぞれの出力を組合せて最終的な出力とした．この時，述語一つあたりに対して，各格約 200 の格フレームと格要素の組合せが出現し，我々の提案手法に比べて計算量は膨大ではあるものの，計算可能な範囲である．この各格に対して独立の SVM を用いて学習を行ったモデルを S0_each とする．

5.1.3 S0′

複数格の同時推定のために我々は単純な解候補削減手法を用意した．この手法では，解析対象述語に近い方から先行する n 個の名詞のみを格要素候補として選ぶ．この時の各格に対する格要素候補数は，提案手法と同程度の格要素の組み合わせ数となるよう調整した値であり，今回は $n = 5$ とした．この単純な解候補削減手法を適用した上で SVM を用いて学習を行ったモデルを S0′ とする．

5.1.4 F0

ベースモデル素性を使用して FNN モデルを実装した．FNN の設計に際しては Matsubayashi and Inui (2017) を参考に，誤差関数にはソフトマックスクロスエントロピーを用い，各隠れ層には batch 正則化 と ReLU 活性化関数を使用した．

5.1.5 F1

格要素候補の分散表現と解析対象述語の分散表現を素性に追加することで，F0 を拡張した．

5.1.6 F2

F1 の述語分散表現を格フレーム内平均ベクトル（MVC）に置き換えた．

5.1.7 F3

F2 に文脈ベクトルとして RNN の出力を追加した．RNN には GRU を使用した．図 1 に F3 の全容を示す．
表 5 にそれぞれの素性組合せを示す．なお，MVP は入力素性として使用していない．

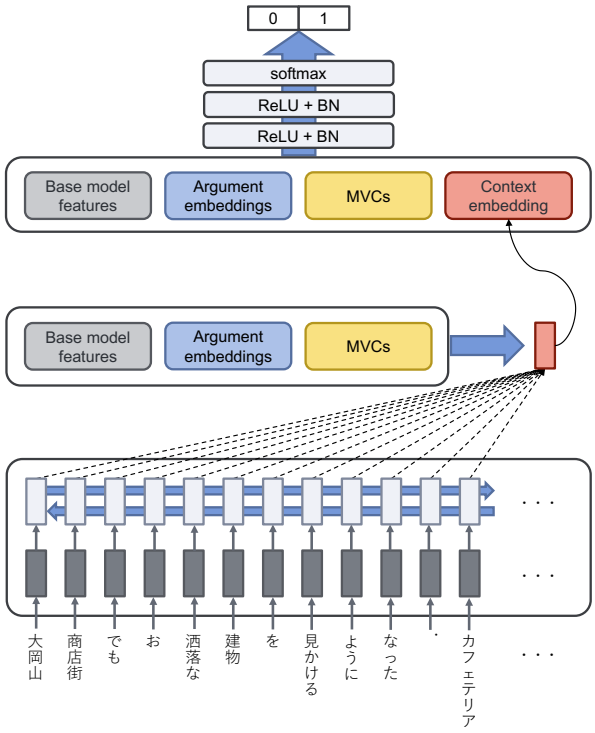

図 1 アテンション付き RNN 追加 FNN モデルのネットワーク構造 (モデル F3)

*6 http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN
*7 http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP
*8 https://taku910.github.io/cabocha/

ⓒ 2018 Information Processing Society of Japan

素性	S0	F0	F1	F2	F3
ベースモデル素性	o	o	o	o	o
格要素分散表現			o	o	o
述語分散表現			o		
MVC				o	o
文脈ベクトル					o

表 5 素性の組合せ

5.2 データセット

実験データとして，『現代日本語書き言葉均衡コーパス』(BCCWJ) (Maekawa et al., 2014) のコアデータ[*9]を使用した．BCCWJ のコアデータ約 2,000 文書に対しては，人手による述語項構造と照応関係が付与されており，これは新聞，雑誌，書籍，白書，Yahoo!知恵袋，Yahoo!ブログの 6 ドメインにまたがっている．ドメインの偏りに注意し，全体の約 4/5 を訓練用データ，約 1/20 を開発用データとし，残りを評価用データとして使用した．複数の格要素が同じ対象を指示している場合 (共参照)，本研究ではコーパスに付与された共参照情報をもとに出力を評価し，正しい照応先と共参照関係にある先行詞のいずれかを対応付けることが出来ていたならば正解とした．本研究で対象とした述語は動詞のみで，形容詞，事態性名詞は扱っていない．

6. 結果と議論

6.1 結果

6.1.1 複数格同時推定の効果

表 6 は BCCWJ におけるゼロ照応解析の実験結果である．S0_each と S0 を比較すると，多くの列において，S0 が S0_each より高い精度を示していることがわかる．ただし文間ガ格，全体ガ格の列においては，S0_each が S0 より高い精度を示している．これは単格の推定では，比較的精度の高いガ格の推定が他の格における誤りから影響を受けないため，複数格同時推定の時より値が良くなっているのだと考えられる．一方で，ニ格，ヲ格については，S0_each は他の格の情報が使えないため比較的精度が低く，全体としての精度も，複数格同時推定を行っている S0 に劣っている．

6.1.2 解候補削減の効果

表 6 で，S0′ と S0 を比較すると，すべての列において，S0 が S0′ より精度が高い．我々はこの結果に対して有意水準 0.1%でマクネマー検定を行い，統計的有意差を確認した．このことから我々の提案した解候補削減手法がうまく機能しているといえる．

6.1.3 分散表現と MVC の効果

ベースラインモデル (F0) に格要素と述語の分散表現を導入すると (F1)，全体の精度が低下した．しかしながら，述語の分散表現を MVC に置き換えることで (F2) 精度は上がり，F0 を上回っている．これは述語の情報 (F1) の代わりに，格フレーム情報を使う (F2) 方がより効果的であることを示している．

6.1.4 文脈ベクトルの効果

ローカルアテンション付き RNN モデルを使用し文脈情報を導入することで (F3)，F2 に比べて改善が見られた．これは，モデルが前方文脈情報を効率的に学習できていることを示唆している．F3 は全体の精度においては S0 に劣っているが，文間照応においては様々な素性を入力としたことにより，S0 より高い精度を示している．

6.2 議論

6.2.1 係り受け関係とゼロ照応解析の精度の関係

我々の提案するモデルは複数の異なる格要素を同時に同定するものである．その効果を見るために，我々はどの格がすでに係り受け解析によって埋まっているかに基づいてテスト事例を分類した．表 7 はそれぞれの組合せにおける精度である．列はすでに直接係り受け関係で埋められた格を示し，行はシステムによって埋められるべき格を示す．例えば，『ガ格』行，『ヲ格』列の数字は，係り受け解析によって『ヲ格』の格要素が与えられた上で，『ガ格』の格要素を同定した精度である．

特に，下線付きのセルが事例数が多いにも関わらず精度が低いため，これらの下線付きセルの述語に関するパフォーマンスを改善することが，ゼロ照応解析全体の精度を上昇させることに対して重要である．

6.2.2 NAIST テキストコーパス (NTC) による実験

我々は NTC を用いることで，本稿の提案手法を Sasano and Kurohashi (2011)， Matsubayashi and Inui (2017) と比較した．表 8 は各タスクの設定を示し，表 9 はそれぞれの実験結果を示す．表 8 で示したように，それぞれの手法のタスク設定は同一でない．従って，厳密な数値の比較は適切ではない．今回の実験では，BCCWJ を訓練データとして使用し，NTC をテストデータとして使用した．NTC を訓練データとして使わなかった理由は，Sasano and Kurohashi (2011) も独自の Web コーパスを訓練データとして使用していたためである（表 8 参照）．また，我々の貢献のひとつが大規模コーパスからの学習を可能とした点であることも理由として挙げられる．

7. 結論

本論文では分散表現で平均化した格フレームによる解候補削減を用いた日本語文内・文間ゼロ照応モデルを提案した．提案した解候補削減アルゴリズムによって大規模な多ドメインコーパスによる訓練を可能とした．また，ローカルアテンション機構付き RNN と FNN を組合せて使用し様々な素性を取り入れることで，文間ゼロ照応解析においてより高い精度が出ることを確認した．

我々の今後の課題はタスクの対象を形容詞，事態性名詞

[*9] http://pj.ninjal.ac.jp/corpus_center/bccwj/

ⓒ 2018 Information Processing Society of Japan

	文内				文間				All			
格	ガ格	ヲ格	ニ格	All	ガ格	ヲ格	ニ格	All	ガ格	ヲ格	ニ格	All
モデル \ 例数	3,137	1,458	873	5,468	2,359	495	243	3,097	5,496	1,953	1,116	8,565
S0_each (ベース)	.570	.730	.757	.643	.085	.016	.144	.080	**.397**	.602	.660	.480
S0′ (ベース)	.490	.712	.725	.589	.032	.016	.140	.038	.331	.584	.632	.435
S0 (ベース)	**.575**	**.758**	**.777**	**.661**	.044	.016	.145	.048	.390	**.628**	**.679**	**.491**
F0 (ベース)	.523	.736	.775	.623	.054	.019	.151	.057	.356	.610	.677	.462
F1 (ベース，格要素，述語)	.470	.682	.762	.564	**.141**	.041	.138	**.126**	.342	.537	.659	.416
F2 (ベース，格要素，MVC)	.563	.707	.773	.641	.103	**.063**	**.154**	.099	.394	.565	.674	.479
F3 (ベース，格要素，MVC，文脈)	.562	.726	.757	.641	.096	.032	.147	.090	.395	.598	.658	.482

表 6 BCCWJ における結果 (F 値)

ゼロ照応として	すでに直接係り受け関係で埋められた格						
埋められるべき格	項なし	ガ格	ヲ格	ニ格	ガ格, ヲ格	ヲ格, ニ格	ガ格, ニ格
外界 or 照応なし	.495 (794)	.817 (3461)	.586 (1011)	.785 (275)	.697 (2046)	.645 (152)	.724 (76)
ガ格	.313 (1645)	-	.282 (1870)	.287 (683)	-	.243 (292)	-
ヲ格	.257 (416)	.384 (656)	-	.247 (81)	-	-	.750 (1222)
ニ格	.505 (111)	.430 (337)	.319 (47)	-	.419 (129)	-	-
ガ格, ヲ格	.112 (492)	-	-	.144 (90)	-	-	-
ヲ格, ニ格	.091 (33)	.057 (35)	-	-	-	-	-
ガ格, ニ格	.228 (281)	-	.189 (122)	-	-	-	-
ガ格, ヲ格, ニ格	.000 (21)	-	-	-	-	-	-

表 7 係り受け関係とゼロ照応解析の関係

にまで拡張し，より実用的なモデルを構築することである．また，対象格要素候補の名詞が先行文脈中の述語にどの格の格要素として取られたかの情報を，我々の提案した解候補削減アルゴリズムに取り入れることで，より良い解候補削減が行えるよう改善する予定である．

謝辞 (Hangyo et al., 2013) に関して詳細な情報をご教示くださった萩行正嗣氏， (Ouchi et al., 2017) の全体像についてご教示くださった大内啓樹氏に厚く御礼申し上げます．

参考文献

Chen Chen and Vincent Ng. 2016. Chinese Zero Pronoun Resolution with Deep Neural Networks. In *ACL*. pages 778–788.

Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. 2012. Building a Diverse Document Leads Corpus Annotated with Semantic Relations. In *PACLIC*. pages 535–544.

Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. 2013. Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions. In *EMNLP*. pages 924–934.

Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What Works and What's Next. In *ACL*. pages 473–483.

Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. 2007. Annotating a Japanese Text Corpus with Predicate-Argument and Coreference Relations. In *Proceedings of the Linguistic Annotation Workshop*. pages 132–139.

Ryu Iida and Massimo Poesio. 2011. A Cross-Lingual ILP Solution to Zero Anaphora Resolution. In *ACL*. pages 804–813.

Ryu Iida, Kentaro Torisawa, Chikara Hashimoto, Jong-Hoon Oh, and Julien Kloetzer. 2015. Intra-sentential Zero Anaphora Resolution using Subject Sharing Recognition. In *EMNLP*. pages 2179–2189.

Ryu Iida, Kentaro Torisawa, Jong-Hoon Oh, Canasai Kruengkrai, and Julien Kloetzer. 2016. Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network. In *EMNLP*. pages 1244–1254.

Kenji Imamura, Kuniko Saito, and Tomoko Izumi. 2009. Discriminative Approach to Predicate-Argument Structure Analysiswith Zero-Anaphora Resolution. In *ACL-IJCNLP*. pages 85–88.

Thorsten Joachims. 2006. Training Linear SVMs in Linear Time. In *Proceedings of the 12th ACM SIGKDD*. pages 217–226.

Daisuke Kawahara and Sadao Kurohashi. 2006. A Fully-

ⓒ 2018 Information Processing Society of Japan

	タスク		訓練コーパス			対象述語		
	文内	文間	新聞	Web	etc.	動詞	形容詞	イベント性名詞
(Matsubayashi and Inui, 2017)	o		o			o	o	o
(Sasano and Kurohashi, 2011)	o	o		o		o	o	
提案手法	o	o	o	o	o	o		

表 8 関連研究のタスク設定

	文内				文間				All			
格	ガ格	ヲ格	ニ格	All	ガ格	ヲ格	ニ格	All	ガ格	ヲ格	ニ格	All
動詞例数	11,559	7,472	4,389	23,420	2,810	229	142	3,181	14,369	7,701	4,531	26,601
S0 (ベース)	.227	.271	.120	.224	.071	.020	.014	.058	.193	.243	.111	.196
(Sasano and Kurohashi, 2011)	.395	.175	.089		.244	.066	.026					
(Matsubayashi and Inui, 2017)	.565	.447	.160	.537								

表 9 NTC を用いた実験結果の F 値

Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis. In *ACL*. pages 176–183.

Taku Kudo, Hiroshi Ichikawa, and Hideto Kazawa. 2014. A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation. In *ACL*. pages 557–562.

Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *EMNLP*. pages 1412–1421.

Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation* 48(2):345–371.

Yuichiroh Matsubayashi and Kentaro Inui. 2017. Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis. In *IJCNLP*. pages 128–133.

Tomas Mikolov, Kai Chen, Grag Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781* .

Hiroki Ouchi, Hiroyuki Shindo, Kevin Duh, and Yuji Matsumoto. 2015. Joint Case Argument Identification for Japanese Predicate Argument Structure Analysis. In *ACL-IJCNLP*. pages 961–970.

Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. 2017. Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis. In *ACL*. pages 1591–1600.

Luz Rello, Ricardo Baeza-Yates, and Ruslan Mitkov. 2012. Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish. In *EACL*. pages 706–715.

Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2008. A Fully-Lexicalized Probabilistic Model for Japanese Zero Anaphora Resolution. In *COLING*. pages 769–776.

Ryohei Sasano and Sadao Kurohashi. 2011. A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames. In *IJCNLP*. pages 758–766.

Tomohide Shibata, Daisuke Kawahara, and Sadao Kurohashi. 2016. Neural Network-Based Model for Japanese Predicate Argument Structure Analysis. In *ACL*. pages 1235–1244.

Kiyoshi Sudo, Satoshi Sekine, and Ralph Grishman. 2001. Automatic pattern acquisition for Japanese information extraction. In *HLT*. pages 1–7.

Hiroaki Yamada, Simone Teufel, and Takenobu Tokunaga. 2017. Designing an annotation scheme for summarizing Japanese judgment documents. In *KSE*. pages 275–280.

Qingyu Yin, Yu Zhang, Weinan Zhang, and Ting Liu. 2017. Chinese Zero Pronoun Resolution with Deep Memory Network. In *EMNLP*. pages 1309–1318.

Jie Zhou and Wei Xu. 2015. End-to-end Learning of Semantic Role Labeling Using Recurrent Neural Networks. In *ACL*. pages 1127–1137.

ⓒ 2018 Information Processing Society of Japan