

事前学習とfinetuningの 類似性に基づくゼロ照応解析

今野颯人¹ 清野舜^{2,1} 松林優一郎^{1,2} 大内啓樹² 乾健太郎^{1,2}

¹ 東北大学 ² 理化学研究所

文の意味理解には省略解析が重要

述語の項の省略

友人が

私は友人を招待した。しかし、来なかった。
後日、来なかった理由を聞こう。

イベント(常識的知識)の省略

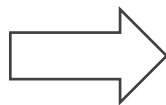
私は友人が
来なかった理由を知らない

述語の項の省略

私が友人に

誰が来なかったのか？

なぜ来なかった理由を聞くのか？



項の省略解析：ゼロ照応解析

述語の項の省略

友人が

私は友人を招待した。しかし、来なかった。
後日、来なかった理由を聞こう。

ゼロ照応

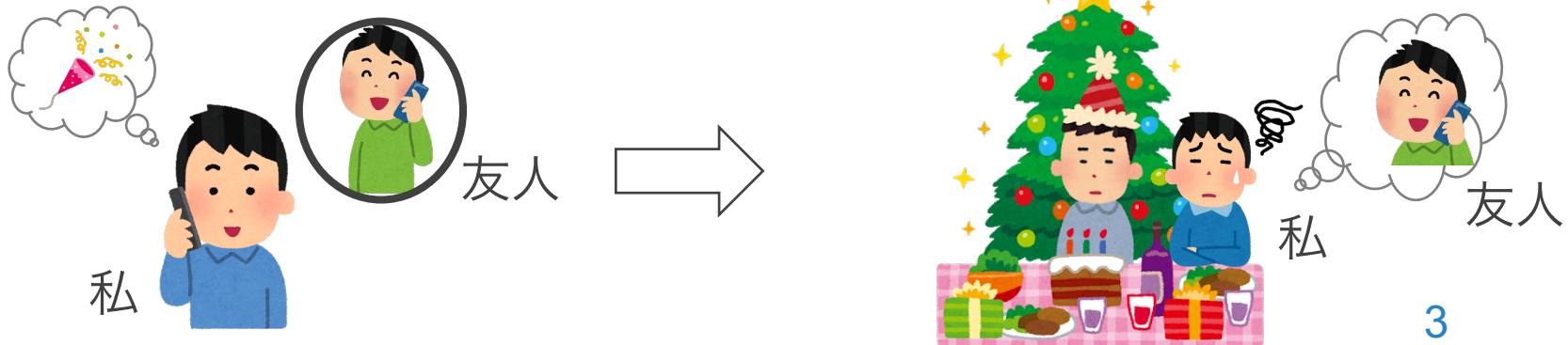
私は友人を招待した。しかし、(φが) 来なかった。

日本語・中国語・スペイン語など

共参照

I invited the friend. But he didn't show up.

英語



ゼロ照応解析には何が必要か？

私は友人を招待した。しかし、(φが)来なかった。



獲得した知識を適用

照応関係の知識を獲得



招待された人

同一人物



来る人

ゼロ照応解析には何が必要か？

省略が起きているテキスト

私は友人を招待した。しかし、(φが)来なかった。



獲得した知識を適用

照応関係の知識を獲得

関係が明示的に書かれているテキスト

同一人物

男が新入社員を招待したが、新入社員は来なかった。

ゼロ照応解析の二つの研究課題

省略が起きているテキスト



私は友人を招待した。しかし、(φが)来なかった。



② どうやって適用する？

① どうやって獲得する？

関係が明示的に書かれているテキスト

同一人物

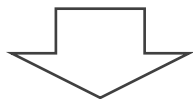
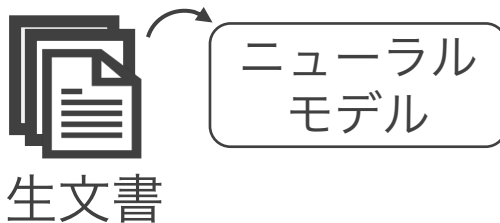


男が新入社員を招待したが、新入社員は来なかった。

本研究における二つのアプローチ

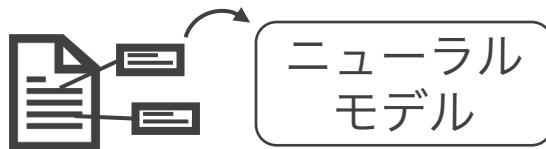
① 知識の獲得：事前学習の工夫

大規模な生文書コーパスからゼロ照応解析に必要な知識を獲得する



② 知識の適用：fine-tuningの工夫

生文書で訓練したモデルをラベル付きデータでさらに学習する



ラベル付きデータ

研究の動向 1/2：大規模な生文書の利用

問題：省略は様々な文脈で出現
少ないラベル付きデータだけでは十分に汎化できない

解決策：大規模な生文書の利用



ゼロ照応解析の
ラベル付きデータ

+



ラベルなしデータ
(生文書)

データが少ない&容易に増やせない

生文書は大量に存在

既存研究で獲得されてきたもの
格フレーム

変化	影響	生活				
要因	が	支障	を	環境	に	及ぼす
⋮		⋮		⋮		

スクリプト知識

「XがYを注文する → XがYを食べる → Xがお金を払う」

選択選好

道路	→	___	が	凍る
店				

研究の動向 1/2：大規模な生文書の利用

■ 格フレーム・選択選好

文脈依存の知識が必要な事例に対処できない

例) 獲得した知識：「受け取る」の主語には「人間」がきやすい

😊 解析可能：太郎は荷物を注文したが、(φが)受け取れなかった。

😞 解析不可：太郎は次郎にボールを投げたが、(φが)受け取れなかった。

■ スクリプト知識

照応先が述語の項ではない事例に対処できない (全体の42%存在※)

※ 松林ら (2015)

例) 獲得した知識：「YがXに投げる」→「Xが受け取る」

😊 解析可能：太郎は次郎にボールを投げたが、(φが)受け取れなかった。

😞 解析不可：次郎ではなく、太郎の作品が最優秀賞だった。
(φが)賞状を受け取った。

研究の動向 2/2 : マスク言語モデルの利用

マスク言語モデル (MLM) ; Devlin et al., (2019)



単語全体の15%をマスクして復元する学習

天気 歩こう

今日は[MASK]が良い日だから、外を[MASK]かな。

ゼロ照応解析に必要な知識を
獲得している？

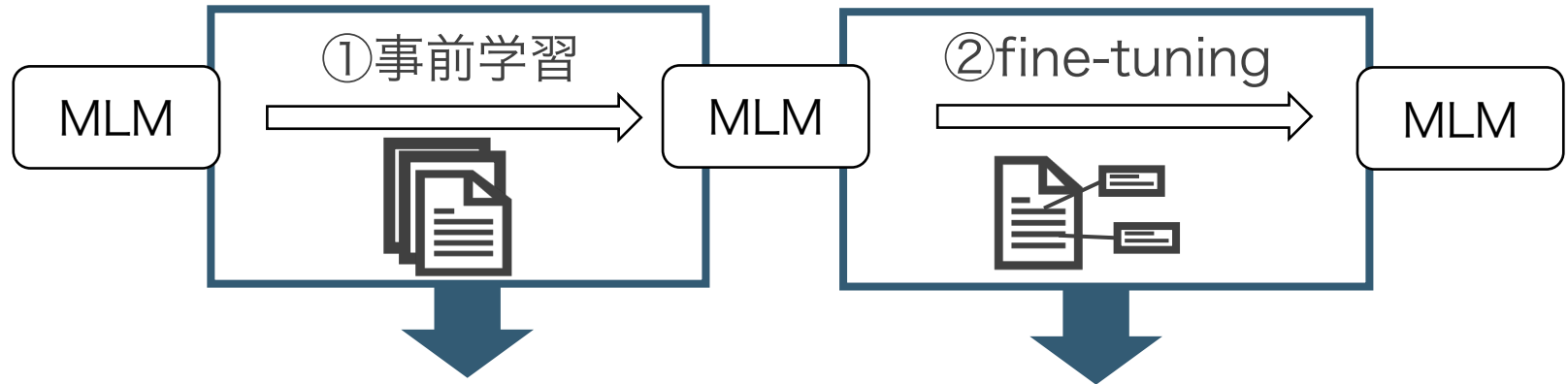
ゼロ照応解析の性能

	F1 Score
M&I'18	55.55
O&K'19	53.50
Konno+'20	64.15

MLMで
性能向上

本研究における二つの提案手法

- ① 知識獲得：事前学習の工夫 ② 知識適用：Fine-tuningの工夫



もっと直接的に
照応関係の知識を獲得する

事前学習との
ギャップを減らす

二つの提案手法を組み合わせることが
日本語のゼロ照応解析に有効かどうかを確かめる

提案手法①：事前学習方法の工夫

従来法

社員

は

男は新入[MASK]を招待したが、新入社員[MASK]来なかった。

1. 全体の単語を15%マスクする
2. マスクされた単語をvocabularyから選ぶ

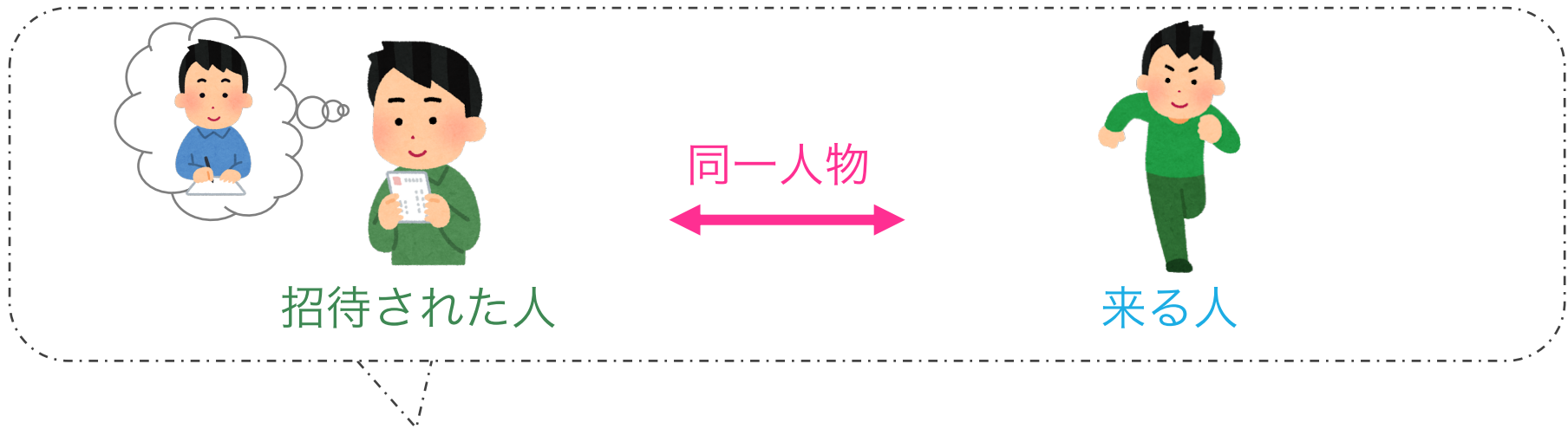
提案法

Predict

男は新入社員を招待したが、[MASK]は来なかった。

1. 生文書から、同じ名詞句が二回以上出現するテキストを抽出
2. そのうち1つをマスク
3. マスクされたところに何が入るのかを文中の単語から選ぶ

提案手法①：事前学習方法の工夫



提案法

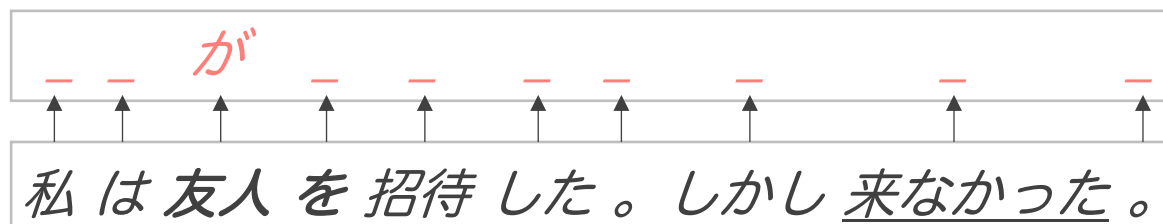
Predict

男は新入社員を招待したが、[MASK] は来なかった。

1. 生文書から、同じ名詞句が二回以上出現するテキストを抽出
 2. そのうち1つをマスク
 3. マスクされたところに何が入るのかを文中の単語から選ぶ
- 直接的に照応関係を学習する
 - 述語の項だけに限らず全ての名詞句が対象

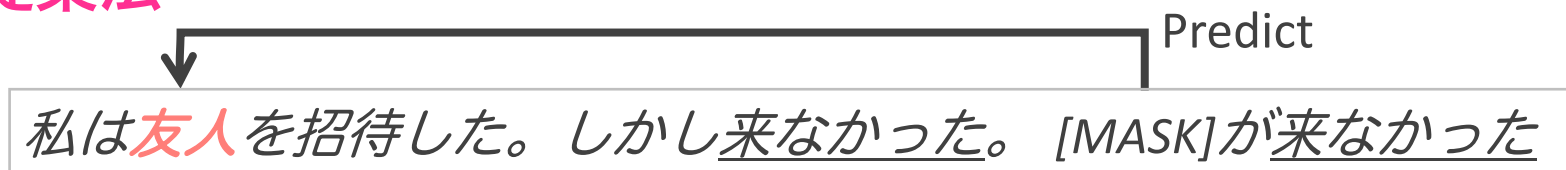
提案手法②：fine-tuningの工夫

従来法



単語ごとに正解ラベルを予測する

提案法



1. 文の後ろに「[MASK] <格助詞> <対象述語>」を付け加える
2. マスクされたところに何が入るのかを文中の単語から選ぶ

提案手法②：fine-tuningの工夫

提案法, 事前学習

事前学習/fine-tuningの
設定を揃えることは効果的
Gururangan et al., 2020;
Yang et al., 2019; ...

Predict

男は新入社員を招待したが、[MASK]は来なかった。

提案法, fine-tuning

Predict

私は友人を招待した。しかし来なかった。[MASK]が来なかった

1. 文の後ろに「[MASK] <格助詞> <対象述語>」を付け加える
2. マスクされたところに何が入るのかを文中の単語から選ぶ

事前学習/fine-tuningの設定を揃えることで
生文書から獲得した知識をうまく省略解析へ適用する

実験における解析対象

データ：NAIST Text Corpus (NTC) 1.5

解析対象：ガ格（主格） / ヲ格（対格） / ニ格（与格）

省略のタイプ：

• Intra (照応先が述語と同じ文内)

友人をパーティーに招待したが、(φが) 来なかった。

• Inter (照応先が述語よりも前の文)

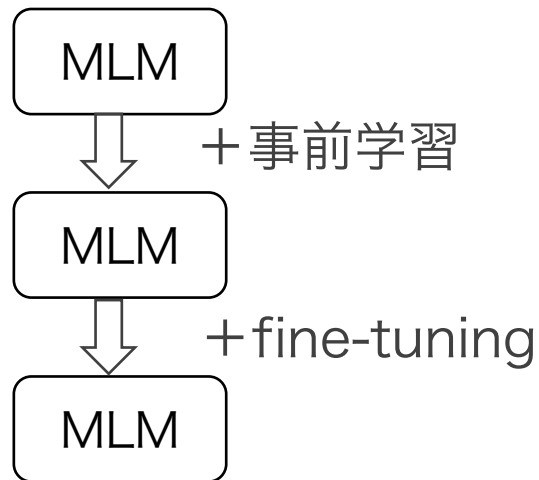
後日、理由を(φが)(φに) 聞こう。

• exophora (照応先が文書内にはない)

話し手（一人称）

実験設定

■ 比較するモデル



	+事前学習	+Fine-tuning
パターン1 (baseline)	なし	従来法
パターン2	なし	提案法
パターン3	従来法	従来法
パターン4	従来法	提案法
パターン5	提案法	従来法
パターン6	提案法	提案法

■ 評価方法

- 評価データ：Naist Text Corpus (NTC) 1.5のTest Set [iida+'10]
- 評価指標： F_1 score

結果 1/2：既存研究とベースラインの比較

設定：入力が1文、解析対象はintraのみ（既存研究と同じ）

モデルサイズ↑
学習更新回数↑



	intra
Ouchi et al., (2017)	47.12
Matsubayashi and Inui (2018)	55.55
Omori and Komachi (2019)	53.50
Konno et al., (2020)	64.15
従来法でのfine-tuningのみ (baseline)	69.32
提案法 でのfine-tuningのみ	69.91

ベースラインの時点で既存研究の性能を上回っている
十分強いベースライン、提案手法でのゲインが信頼できる

結果 2/2 : 提案手法の効果を検証

設定 : 入力が複数文、解析対象はintra, inter, exophora

		事例数:	(14,066)	(6,159)	(4,081)	(3,826)
		事前学習	All	intra	inter	exo
入力が1文:		従来法	-	69.32	-	-
		提案法	-	69.91	-	-
入力が複数文:		従来法	62.27±0.43	71.55	44.30	64.04
		提案法	62.47±0.53	71.09	45.20	64.41
+ 事前学習		従来法	62.54±0.47	71.82	44.98	63.94
		従来法	62.85±0.19	71.52	45.97	64.55
		提案法	63.06±0.19	71.96	46.37	64.42
		提案法	64.18±0.23	72.67	48.41	65.40

- ① 周辺文脈を追加することでintraの性能が向上
- ② 事前学習によって全てのカテゴリの性能が向上

結果 2/2 : 提案手法の効果を検証

設定 : 入力が複数文、解析対象はintra, inter, exophora

		事例数:	(14,066)	(6,159)	(4,081)	(3,826)
事前学習	Fine-tuning		All	intra	inter	exo
	従来法		-	69.32	-	-
	提案法		-	69.91	-	-
	従来法		62.27±0.43	71.55	44.30	64.04
	提案法		62.47±0.53	71.09	45.20	64.41
	従来法	従来法	62.54±0.47	71.82	44.98	63.94
	従来法	提案法	62.85±0.19	71.52	45.97	64.55
	提案法	従来法	63.06±0.19	71.96	46.37	64.42
	提案法	提案法	64.18±0.23	72.67	48.41	65.40

知識獲得の工夫



直接的に照応関係の知識を獲得する方法は効果的

結果 2/2 : 提案手法の効果を検証

設定 : 入力が複数文、解析対象はintra, inter, exophora

		事例数:	(14,066)	(6,159)	(4,081)	(3,826)
事前学習	Fine-tuning		All	intra	inter	exo
	従来法		-	69.32	-	-
	提案法		-	69.91	-	-
	従来法		62.27±0.43	71.55	44.30	64.04
	提案法		62.47±0.53	71.09	45.20	64.41
従来法	従来法		62.54±0.47	71.82	44.98	63.94
従来法	提案法		62.85±0.19	71.52	45.97	64.55
提案法	従来法		63.06±0.19	71.96	46.37	64.42
提案法	提案法		64.18±0.23	72.67	48.41	65.40

知識適用の工夫



事前学習とfine-tuningのギャップを緩和することが効果的

世界最高性能を達成
しかしinterはまだ低い

考察：事例のタイプ別の性能

事例の種類別におけるInterのRecall

事例の種類	従来法-従来法	提案法-提案法	事例数
省略された項がテキストに一回のみ出現	35.69	➡ 39.57	1218
省略された項がテキストに二回以上出現	53.10	54.13	872
省略された項が述語の1文前に出現	48.86	➡ 51.96	
省略された項が述語の2文前に出現	37.71	➡ 42.58	
省略された項が述語の2文より前に出現	40.50	40.31	
述語が受動態	25.24	➡ 29.13	206
述語が能動態	45.02	47.42	1877
All	43.11	45.65	2090

緑：3ポイント以上 向上

- 解析が難しい(複雑な照応関係にある)事例が解けるように
- 省略された項が述語から遠すぎると効果なし
提案手法の事前学習：表層形が同じ名詞句を当てる
➡ 同じ表層形でも距離が遠いと違う概念を指している可能性が高い

考察：事例のタイプ別の性能

事例の種類別におけるInterのRecall

事例の種類	従来法-従来法	提案法-提案法	事例数
省略された項がテキストに一回のみ出現	35.69	39.57	1218
省略された項がテキストに二回以上出現	53.10	54.13	872
省略された項が述語の1文前に出現	48.86	51.96	
省略された項が述語の2文前に出現	37.71	42.58	
省略された項が述語の2文より前に出現	40.50	40.31	
述語が受動態	25.24	29.13	206
述語が能動態	45.02	47.42	1877
All	43.11	45.65	2090

- 述語が受動態だと性能が低い
 - 格交代が起きる（例：AはBに食べられる → BがAを食べる）
 - NTCのアノテーションは全て**能動態**
 - 生文書から格交代を学習するのは困難

結論

- タスク：日本語の項の省略解析 (ゼロ照応解析)
- 二つの手法を提案
 - ① 照応関係の知識を獲得するための事前学習方法
 - ② 獲得した知識をうまく解析へ適用するためのfine-tuning方法
- 提案手法による結果：
 - 特に難しい事例で性能が向上
 - 述語と省略された項が大きく離れた事例では効果がみられない
- 今後の展望：
 - 述語から遠い位置にある名詞句との関係を捉えるための工夫
 - 例) 文書中で主題となっている名詞句は省略されやすい
 - 受動態でアノテーションされている別のデータセットでも提案手法が効果的であるのかを検証する