

日本語述語項構造解析のための自己回帰モデル

宮脇 峻平^{1,*}, 加藤 拓真¹, 今野 颯人¹, 大内 啓樹^{2,1},
清野 舜^{2,1}, 松林 優一郎^{1,2}, 高橋 諒^{1,2}, 乾 健太郎^{1,2}

* miyawaki.shumpei@ecei.tohoku.ac.jp ¹ 東北大学 ² 理化学研究所

概要

日本語述語項構造解析のための自己回帰モデルを提案

モデルの予測結果から、高い確信度の情報を用いて解き直し、項の格関係における予測精度の向上に取り組む

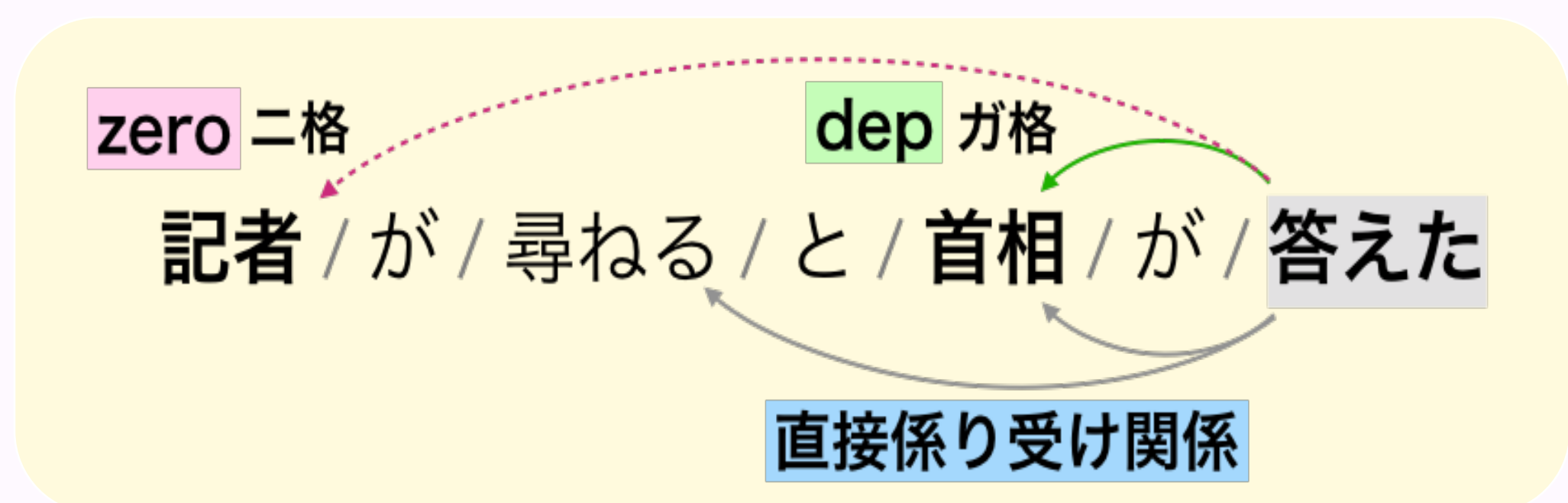
ベースモデルとの実験結果を比較・分析

ベースラインモデル [Matsubayashi+'18] に自己回帰を組み合わせることで、解析精度の向上を達成

提案手法の予測結果から、自己回帰の効果を定量的に分析

背景

述語項構造解析 では、項の種類 によって 予測の難易度 が異なる



項別予測精度		
F1	dep	zero
85.25	91.11	57.59

[Matsubayashi+'18]

自己回帰 の枠組みで

簡単に予測できる項 の情報を

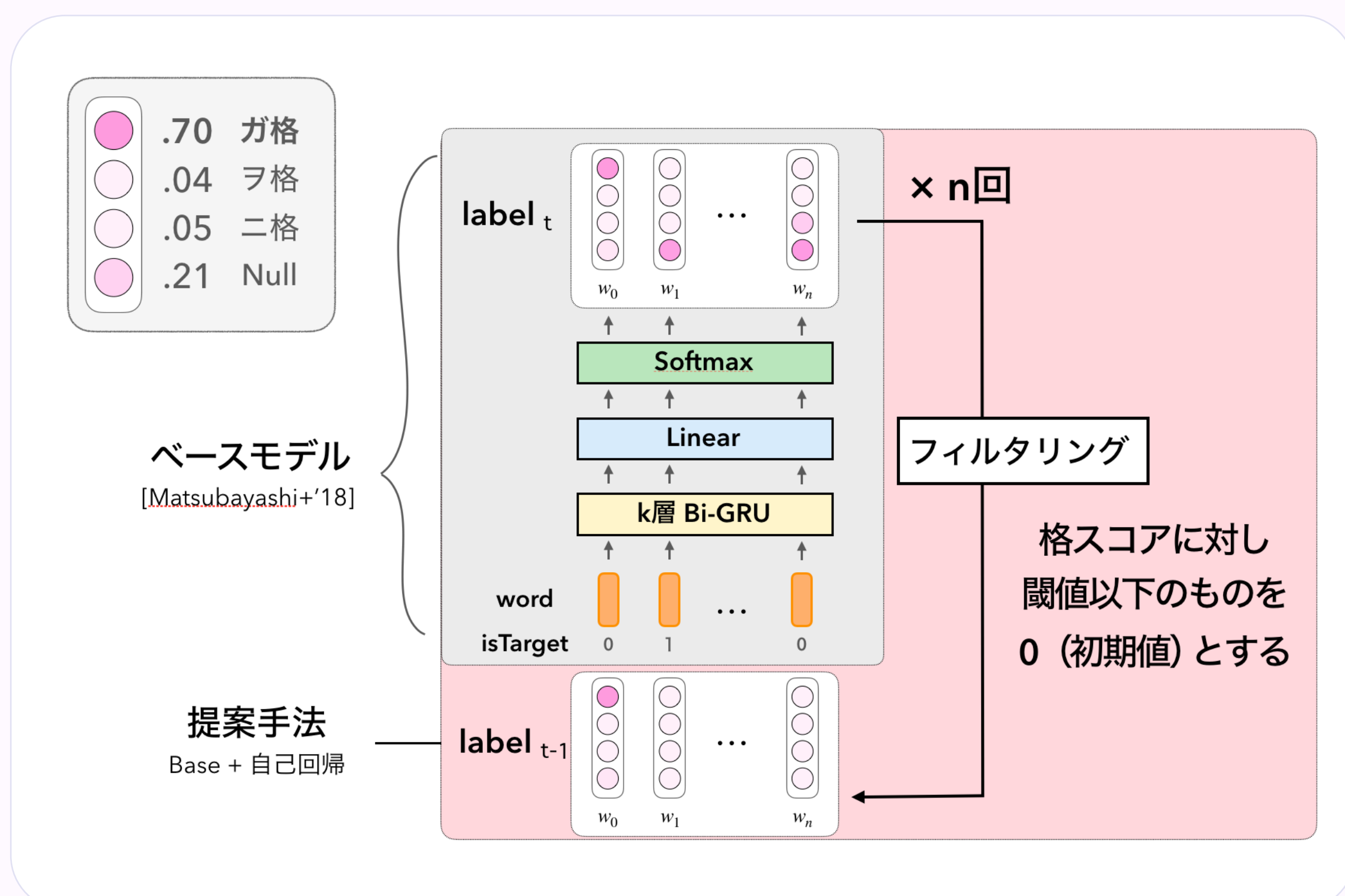
予測困難な項の予測 に活用したい

提案手法

提案手法

- 出力値（各ラベルのスコア）を入力値に連結したのち、再予測を行う
- その際、閾値以上（高い確信度）のものを次の入力に使用する

閾値を設けることで、
確信度の低い予測結果が
自己回帰のノイズになることを防ぐ



実験

実験設定

Embedding	256次元
Bi-GRU	10層, 256次元
Optimizer	Adam
Dropout	0.1
Seed	3種類

データセット

NAIST Text Corpus

	文数	述語数
訓練	23,965	68,248
開発	4,772	13,731
評価	9,114	26,367

実験結果・分析

ベースモデル vs 提案モデル

	F1	適合率	再現率	DEP	ZERO
BASE	83.42	88.27	79.08	90.17	54.07
Matsubayashi+(2018) BASE	83.09	85.64	80.70	89.76	53.26
Omori +(2019) BASE	83.62	—	—	90.09	51.87
BASE + 自己回帰	83.75	87.92	79.96	90.35	55.22
差分 (- BASE)	+0.33	-0.35	+0.88	+0.18	+1.15
Matsubayashi+ (2018) MP-POOL-ATT	83.62	86.09	81.29	90.07	54.53
Omori+ (2019) Multi-ALL	83.82	—	—	90.15	53.50

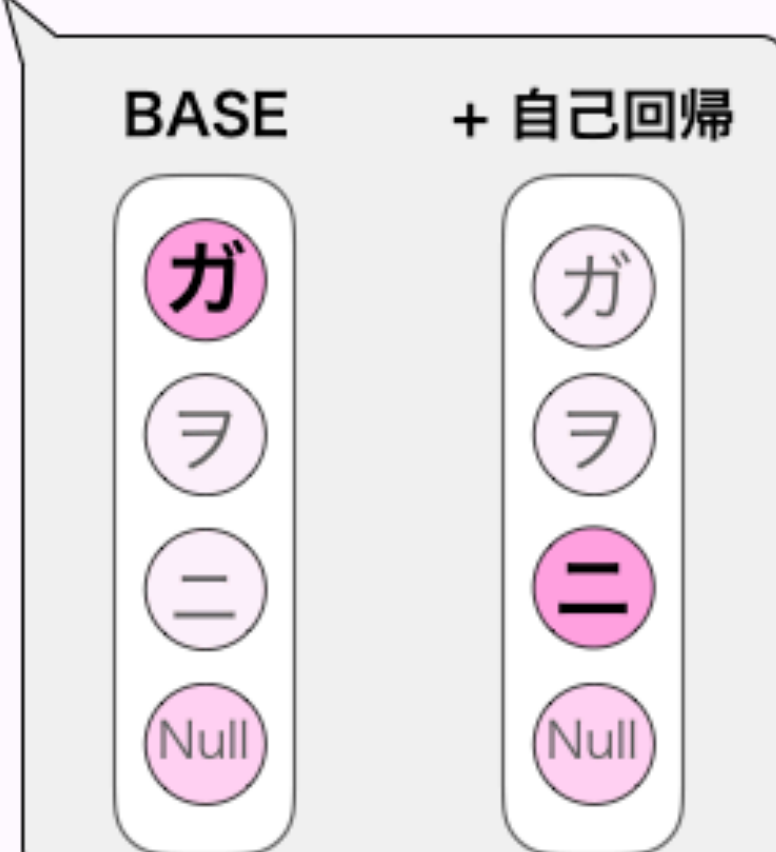
ベースラインとの実験結果の比較

分析: 自己回帰によって正しい予測をした例

会場 / の / 施設 / を / 含め / 難問 / が / 多い

	BASE			BASE + 自己回帰		
	ALL	DEP	ZERO	ALL	DEP	ZERO
正解の予測数	25,576	22,632	2,919	25,860	22,817	3,015
正解の非予測数	6,764	3,271	3,240	6,480	3,086	3,144
不正解の予測数	3,400	1,665	1,719	3,552	1,786	1,745

自己回帰による正解数の影響



- 提案手法において、いずれも F値 が同等以上
- また再現率の向上より、自己回帰を行うことで、ベースが予測できない項の予測が可能になった
- 直接係り受け（DEP）に比べ、ゼロ照応（ZERO）における再現率・F値の上がり幅が特に大きくなった

今後の課題

- 複数述語を考慮した自己回帰モデルの提案
 - 異なる述語での同一項の共有
 - 述語間での依存関係の考慮
- 適切なハイパーパラメータ・手法の探索（フィルタリング・再入力値）
- 自己回帰による影響・詳細なエラー分析