Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis

Yuichiroh Matsubayashi♠ and Kentaro Inui♠♦
♠Tohoku University, Japan, ♦RIKEN, Japan
{y-matsu, inui}@ecei.tohoku.ac.jp

Overview

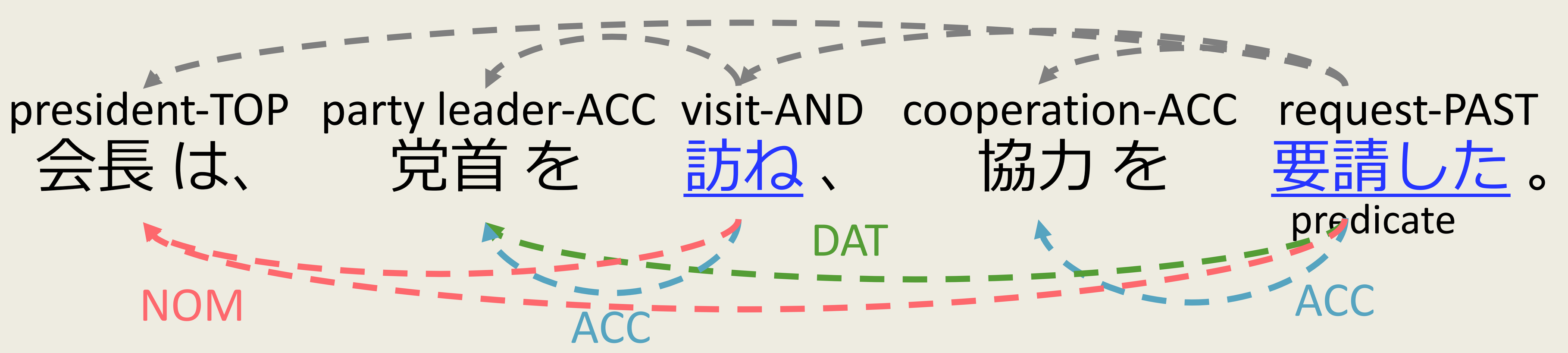
- We show the importance of designing a sophisticated local model before exploring global solution algorithms in Japanese predicate-argument structure (PAS) analysis by demonstrating its impact on the overall performance through an extensive empirical evaluation.
- A local model alone is able to significantly outperform the state-of-the-art global models by incorporating a broad range of local features proposed in the literature and training a neural network for combining them.
- Our best local model achieved 13% error reduction in F1 compared with the state of the art.
- Global models are expected to improve the performance by incorporating such a strong local model.

Motivation

- The research trend in Japanese PAS analysis is shifting from pointwise prediction models with local features to global models designed to search for globally optimal solutions.
- However, the existing global models tend to employ only relatively simple local features; therefore, the overall performance gains are rather limited.

Task Setting

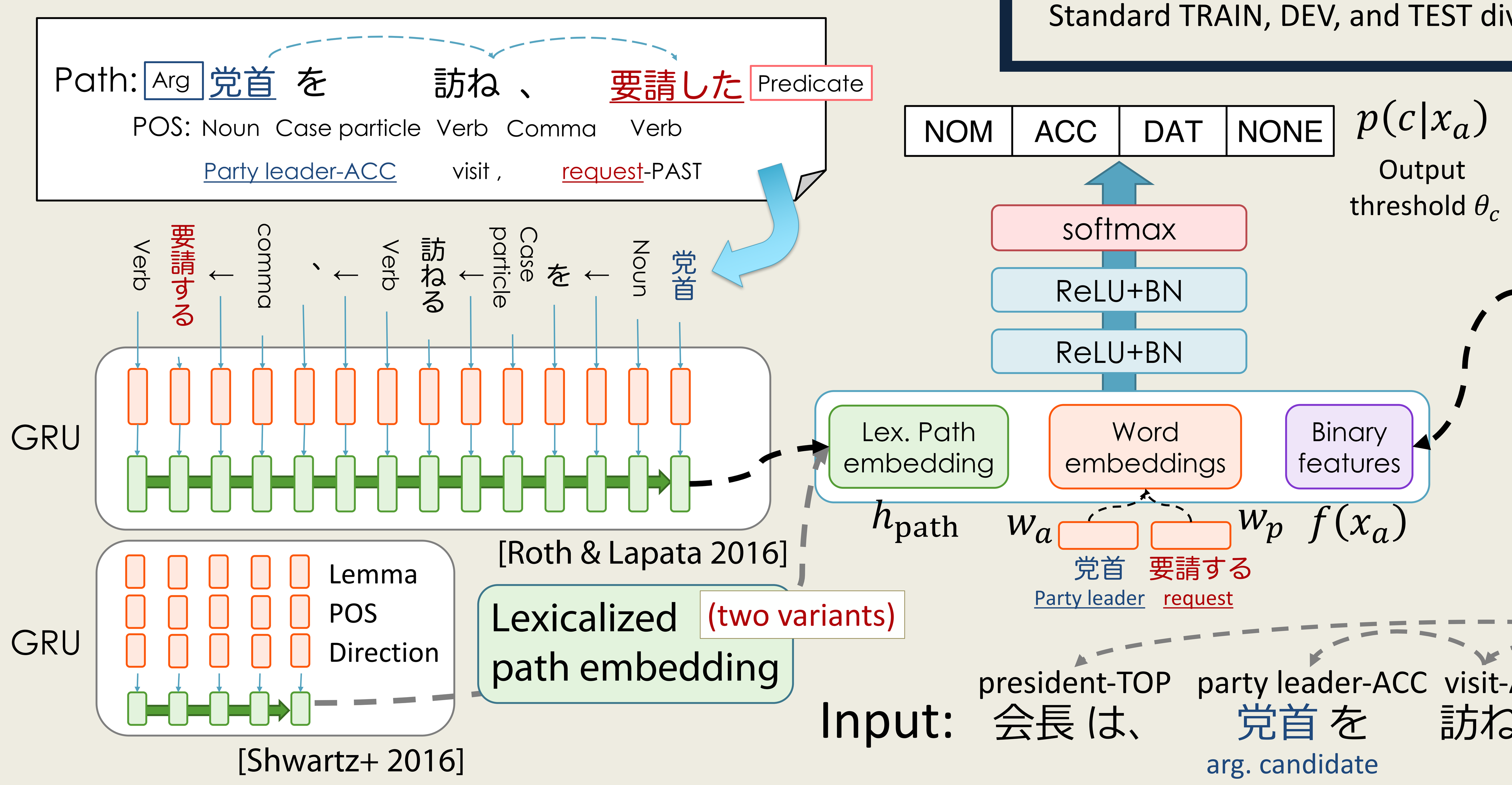
Input: tokenized sentence and predicate positions



Output: NOM, ACC, DAT arguments for each predicate in the sentence

Dataset: NAIST text corpus 1.5 [Iida+, 2007]
Standard TRAIN, DEV, and TEST divisions [Taira+, 2008]

Model



For predicate *p*	surface, lemma, POS, type of conjugated form, nominal form of nominal verb, voice suffixes (-reru, -seru, -dekiru, -tearu)
For argument candidate *a*	surface, lemma, POS, NE tag, whether *a* is head of *bunsetsu*, particles in *bunsetsu*, right neighbor token's lemma and POS
Between predicate and argument candidate	case markers of other dependents of *p*, whether *a* precedes *p*, whether *a* and *p* are in the same *bunsetsu*, token- and dependency-based distances, naive dependency path sequence

binary features

Model	Binary feats.	All: F_1 (σ)	All: Prec.	All: Rec.	F_1 in different dependency distance: *Dep*	*Zero*	2	3	4	≥ 5
B	all	82.02 ($\pm$0.13)	83.45	80.64	89.11	49.59	57.97	47.2	37	21
B	−cases	81.64 ($\pm$0.19)	83.88	79.52	88.77	48.04	56.60	45.0	36	21
WB	all	82.40 ($\pm$0.20)	85.30	79.70	89.26	49.93	58.14	47.4	36	23
WBP-Roth	all	82.43 ($\pm$0.15)	84.87	80.13	89.46	50.89	58.63	49.4	39	**24**
WBP-Shwartz	all	83.26 ($\pm$0.13)	85.51	81.13	89.88	51.86	60.29	49.0	39	22
WBP-Shwartz	−word	83.23 ($\pm$0.11)	85.77	80.84	89.82	51.76	60.33	49.3	38	21
WBP-Shwartz	−{word, path}	83.28 ($\pm$0.16)	85.77	80.93	89.89	51.79	60.17	49.4	38	23
WBP-Shwartz (ens)	−{word, path}	**83.85**	**85.87**	**81.93**	**90.24**	**53.66**	**61.94**	**51.8**	**40**	**24**
WBP-Roth	−{word, path}	82.26 ($\pm$0.12)	84.77	79.90	89.28	50.15	57.72	49.1	38	**24**
BP-Roth	−{word, path}	82.03 ($\pm$0.19)	84.02	80.14	89.07	49.04	57.56	46.9	34	18
WB	−{word, path}	82.05 ($\pm$0.19)	85.42	78.95	89.18	47.21	55.42	43.9	34	21
B	−{word, path}	78.54 ($\pm$0.12)	79.48	77.63	85.59	40.97	49.96	36.8	22	9.1

W: word embedding, B: binary features, P: path embedding
P-Roth: method of [Roth & Lapata, 2016], P-Shwartz: method of [Shwartz et al., 2016]

Model	ALL	Dep: ALL	Dep: NOM	Dep: ACC	Dep: DAT	Zero: ALL	Zero: NOM	Zero: ACC	Zero: DAT
On NTC 1.5									
WBP-Shwartz (ens) −{word, path}	**83.85**	**90.24**	**91.59**	**95.29**	62.61	**53.66**	**56.47**	**44.7**	**16**
B	82.02	89.11	90.45	94.61	60.91	49.59	52.73	38.3	11
(Ouchi et al., 2015)-local	78.15	85.06	86.50	92.84	30.97	41.65	45.56	21.4	0.8
(Ouchi et al., 2015)-global	79.23	86.07	88.13	92.74	38.39	44.09	48.11	24.4	4.8
(Ouchi et al., 2017)-multi-seq	81.42	88.17	88.75	93.68	**64.38**	47.12	50.65	32.4	7.5
Subject anaphora resolution on modified NTC, cited from (Iida et al., 2016)									
(Ouchi et al., 2015)-global							**57.3**		
(Iida et al., 2015)							41.1		
(Iida et al., 2016)							52.5		

Note that [Ouchi et al., 2017] does not use preprocessed syntactic dependency

Impact of Feature Representations

- The *case markers of the other dependents feature* significantly improves the prediction in both Dep and Zero cases, especially on Zero argument detection.
- WBP-Roth and WB compete in our setting
  - The word inputs at both ends of the path embedding overlap with the word embedding and the additional effect of the path embedding is rather limited.
- WBP-Shwartz obtains better result compared with WBP-Roth
- The performance of WBP-Shwartz remains without lexical and path binary features.

Comparison to Related Work

- B model that uses only binary features already outperforms the state-of-the-art global models [Ouchi+, 2015, 2017]
  - [Ouchi et al. 2015] contains almost the same binary features as ours.
- The WBP-Shwartz (ens) shows a further 1.8 points improvement in overall F1, which achieves 13% error reduction compared with the state-of-the-art global model.
  - 81.42% of [Ouchi et al., 2017]-multi-seq