Dependency Parsing as Sequence Labeling with Head-Based Encoding and Multi-Task Learning

Ophélie Lacroix

Siteimprove, Copenhagen, Denmark
ola@siteimprove.com

August 27, 2019

Dependency Parsing as Sequence Labeling

1. Encoding the trees into sequences of labels
2. Using a sequence tagger to learn
   and predict the labels
3. Decoding the predicted labels to build the trees

Alternative to transition-based and graph-based approaches

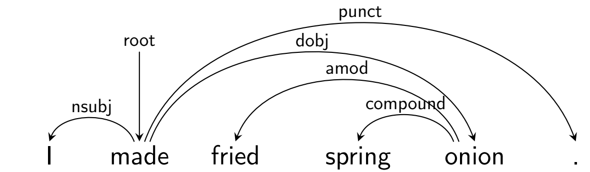
Recent studies : [Strzyz et al., 2019]

- good speed-accuracy trade-off
- compare several encodings
- best encoding relies on Part-of-Speech (PoS) tags

Dependency Tree as Sequence of Labels

Relative PoS-based (RPT) encoding of the dependencies [Strzyz et al., 2019] inspired by [Spoustová and Spousta, 2010]

- *what is the PoS-tag of the head ?*
- *what is its relative position to the child ?*



PRON	VERB	VERB	NOUN	NOUN	PUNCT	PoS
VERB^{+1}	ROOT	NOUN^{+2}	NOUN^{+1}	VERB^{-2}	VERB^{-2}	RPT
nsubj	root	amod	compound	dobj	punct	LABELS

Some flaws

- PoS-tagging is a necessary pre-processing task for RPT
- [Strzyz et al., 2019] no evaluation of PoS-tagging speed

⚠ Neural transition-based parsers can leave-out PoS-tags

→ multi-task learning of PoS-tagging and dependency parsing

⚠ Rare and ambiguous PoS-tags are not reliable

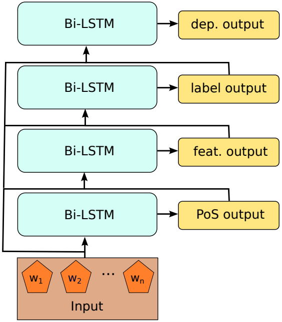
→ new head-based encoding

Sequence Labeling Pipeline : PoS-tagging and Dependency Parsing

Multi-task learning strategies

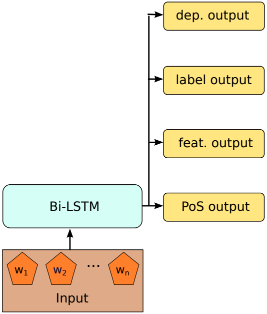
Stacked
[Hashimoto et al., 2017]

one layer = one task



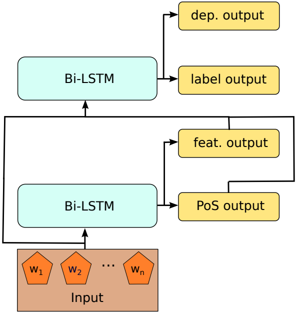
Shared
[Søgaard and Goldberg, 2016]

share parameters



Combined Multi-task Learning Strategy

Combined = Shared + Stacked



Experiments: Multi-task Learning Strategies

	Relative PoS-tag based dep. encoding					
	Shared		Stacked		Combined	
Lang.	**UAS**	**LAS**	**UAS**	**LAS**	**UAS**	**LAS**
cs	85.36	81.29	**87.50**†	**83.66**†	86.84	82.92
en	80.33	76.17	**82.50**	**78.41**	81.88	77.87
fi	77.05	71.37	**80.80**†	**75.95**†	79.85	74.85
grc	67.98	60.28	68.61	61.29	**68.96**	**61.41**
he	72.28	65.52	**77.80**†	**71.56**†	75.53	69.27
kk	42.89	18.88	41.27	17.36	**44.08**†	**19.36**†
ta	62.89	50.65	63.11	51.37	**63.45**	**52.29**†
zh	68.28	61.90	70.91	64.66	**71.00**	**65.00**
avg	69.63	60.76	**71.56**	**63.03**	71.45	62.87

Combined strategy:

- parsing speed increased by 48% compared to the **Stacked**

A New Encoding?

Flaws of the *relative PoS-tag based* encoding:

- infrequent tags:
  - 90% tokens (in En UD) are tagged with the same 15 Rpt tags among 198!
- consecutive PoS-tags with similar roles:
  - Noun & Propn or Verb & Aux
  - make the prediction of the relative position less accurate

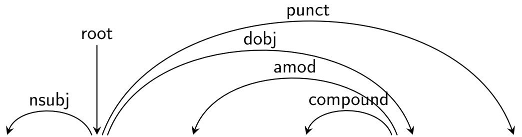
New encoding:
Relative Head-Based Encoding

- *head*-tags instead of PoS-tags
- reduces the size of the tagset

Relative Head-Based Encoding

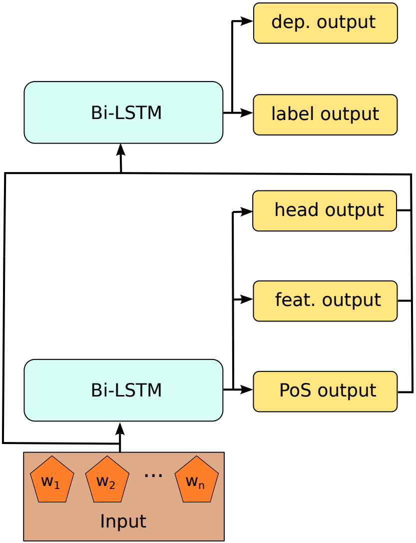
Coarse-grained VS fine-grained encoding strategies

- Relative Unique Head (RUH): X
- Relative Chunk Head (RCH): VP, NP, AP, X



I	made	fried	spring	onions	.	
PRON	VERB	VERB	NOUN	NOUN	PUNCT	PoS
VERB^{+1}	ROOT	NOUN^{+2}	NOUN^{+1}	VERB^{-2}	VERB^{-2}	RPT
	X			X		U.Head
X^{+1}	ROOT	X^{+1}	X^{+1}	X^{-1}	X^{-1}	RUH
	VP			NP		C.Head
VP^{+1}	ROOT	NP^{+1}	NP^{+1}	VP^{-1}	VP^{-1}	RCH

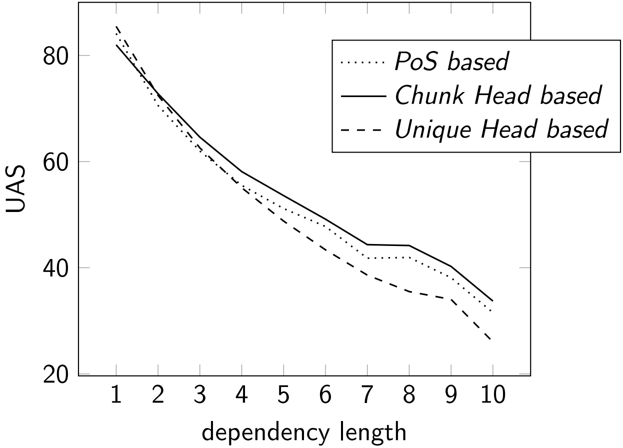
Combined Strategy with Head Based Encoding



Experiments: Encodings Comparison

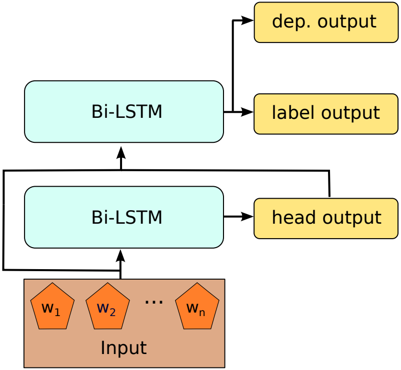
	Rel. PoS-Tag based encoding		Rel. Unique Head based encoding		Relative Chunk Head based encoding	
Lang.	**UAS**	**LAS**	**UAS**	**LAS**	**UAS**	**LAS**
cs	**86.84**†	82.92	86.24	**83.11**	86.09	82.31
en	81.88	77.87	81.48	77.34	**82.70**†	**78.76**†
fi	79.85	74.85	77.33	72.36	**79.89**	**75.08**
grc	**68.96**	**61.41**	67.61	59.72	68.71	61.39
he	75.53	69.27	**81.48**†	**74.12**†	76.93	70.13
kk	44.08	19.36	**47.61**†	**21.70**†	40.19	18.95
ta	63.45	52.29	62.13	50.52	**65.48**†	**54.32**†
zh	71.00	65.00	71.85	65.26	**73.02**†	**66.82**†
avg.	71.45	62.87	**71.97**	63.02	71.63	**63.47**

Dependency Length



- with RUH : many infrequent high relative position
- precision on heads : -6 on chunk heads compared to PoS-tags

Ablating PoS-tagging



Experiments: Ablating PoS tagging

	Relative Chunk Head based encoding			
			-PoS/feat	
Lang.	**UAS**	**LAS**	**UAS**	**LAS**
cs	**86.09**	**82.31**	85.96	82.06
en	**82.70**	**78.76**	81.61	77.33
fi	**79.89**	**75.08**	78.43	72.64
grc	**68.71**	**61.39**	67.91	60.44
he	76.93	**70.13**	**77.49**	69.97
kk	**40.19**	**18.95**	37.30	17.04
ta	**65.48**	**54.32**	60.70	49.04
zh	**73.02**	**66.82**	71.17	64.34
avg.	**71.63**	**63.47**	70.07	61.61

Conclusion

- Multi-task learning **combined** strategy
  - on par with a sequential (stacked) approach
  - significantly faster at parsing sentences
- New **head-based encoding** of the dependencies as labels
  - outperforms the *PoS-based* encoding for a majority of the languages
  - choice of the head tagset is crucial

Hashimoto, K., Xiong, C., Tsuruoka, Y., and Socher, R. (2017).
A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks.
In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*.

Søgaard, A. and Goldberg, Y. (2016).
Deep multi-task learning with low level tasks supervised at lower layers.
In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*.

Spoustová, D. and Spousta, M. (2010).
Dependency Parsing as a Sequence Labeling Task.
The Prague Bulletin of Mathematical Linguistics.

Strzyz, M., Vilares, D., and Gómez-Rodríguez, C. (2019).
Viable Dependency Parsing as Sequence Labeling.

In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)*.