Morphological Inflection Generation with Hard Monotonic Attention

Presenter: Baosong Yang

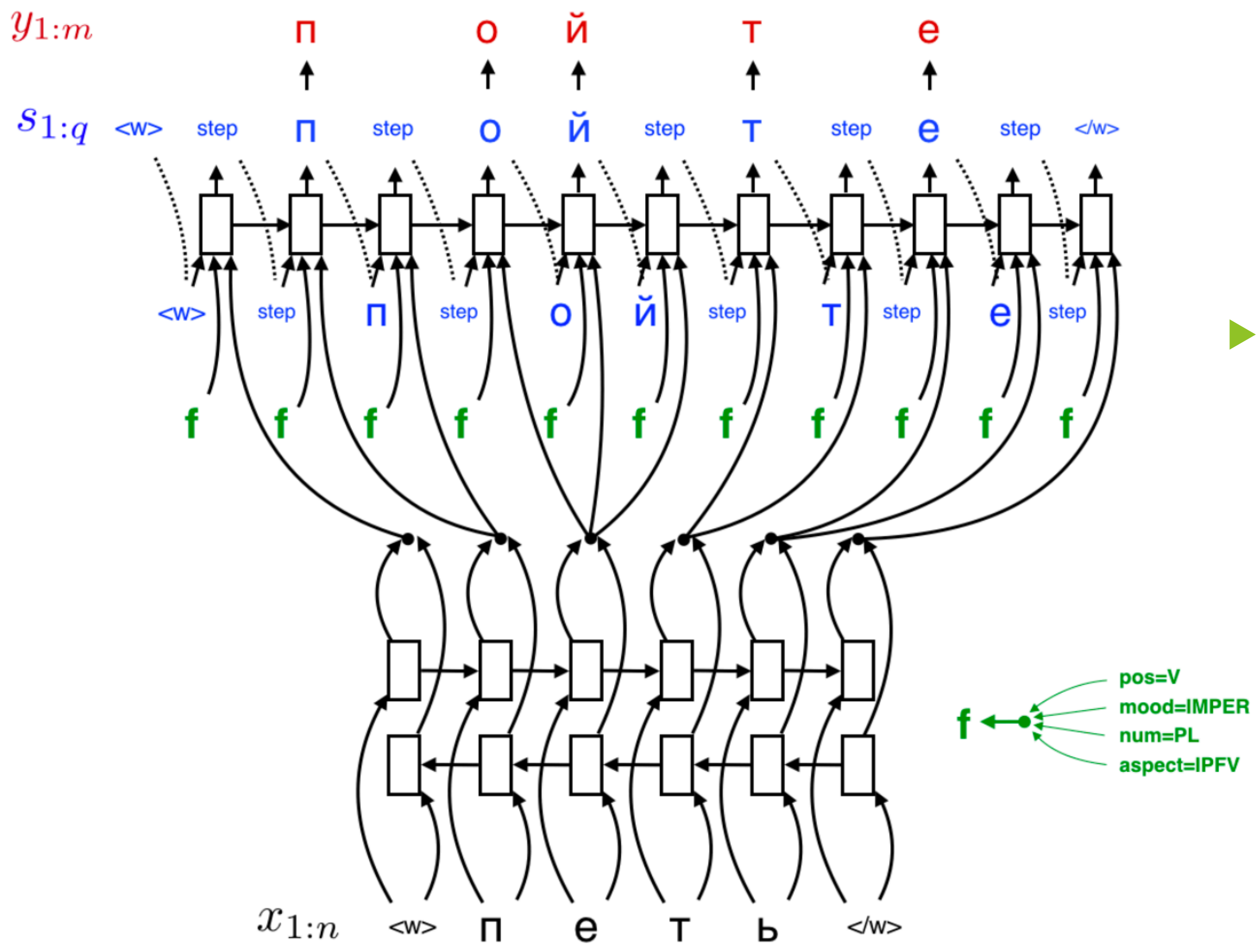
Contents

- Motivation
- Model
- Experiments
- Conclusion

Motivation

- Morphological inflection generation.
  - hard + {POS = adjective, gender = masculine, type = superlative} => hardest.
  - Dealing with data sparsity in morphologically rich languages.
- The neural sequence-to-sequence models require large training sets.
- A hard attention model for nearly- monotonic sequence to sequence learning.

Model



- The attention is promoted to the next input element once a step action is predicted.

Training the Model

- Learning Hard Alignments using character alignment model.
- Deriving Oracle Actions

Experiments

	13SIA	2PIE	2PKE	rP	Avg.
MED (Kann and Schütze, 2016a)	83.9	95	87.6	84	87.62
NWFST (Rastogi et al., 2016)	86.8	94.8	87.9	81.1	87.65
LAT (Dreyer et al., 2008)	**87.5**	93.4	87.4	84.9	88.3
Soft	83.1	93.8	88	83.2	87
Hard	85.8	**95.1**	**89.5**	**87.2**	**89.44**

Table 1: Results on the CELEX dataset

	DE-N	DE-V	ES-V	FI-NA	FI-V	FR-V	NL-V	Avg.
Durrett and DeNero (2013)	88.31	94.76	99.61	92.14	97.23	98.80	90.50	94.47
Nicolai et al. (2015)	88.6	97.50	99.80	93.00	**98.10**	**99.20**	96.10	96.04
Faruqui et al. (2016)	88.12	**97.72**	**99.81**	95.44	97.81	98.82	96.71	96.34
Yu et al. (2016)	87.5	92.11	99.52	95.48	**98.10**	98.65	95.90	95.32
Soft	88.18	95.62	99.73	93.16	97.74	98.79	96.73	95.7
Hard	**88.87**	97.35	99.79	**95.75**	98.07	99.04	**97.03**	**96.55**

Table 2: Results on the Wiktionary datasets

	suffixing+stem changes			circ.	suffixing+agg.+v.h.			c.h.	templatic		
	RU	DE	ES	GE	FI	TU	HU	NA	AR	MA	Avg.
MED	91.46	95.8	98.84	98.5	95.47	98.93	96.8	91.48	**99.3**	**88.99**	95.56
Soft	92.18	96.51	98.88	**98.88**	**96.99**	**99.37**	**97.01**	**95.41**	**99.3**	88.86	**96.34**
Hard	**92.21**	**96.58**	**98.92**	98.12	95.91	97.99	96.25	93.01	98.77	88.32	95.61

- Very small (500 training samples).
- 360k training examples per language
- 12,800 training and 1600 development examples per language.

Conclusion

- Hard Attention: Jointly align and transduce while using a focused representation at each step, rather then the weighted sum of representations used in the soft attention model.
- Soft attention mechanism insufficiently learning enough information from small training set.
- May beneficial for morphological language translation (e.g. German). Replace Byte-pair-encoding