Comparing attention mechanisms for Hungarian morphological analysis

Introduction

- Morphology: study of word formation
- Hungarian morphology is highly complex
  - agglutinative (inflection goes at the end of the word)
  - vowel harmony
  - no grammatical gender
  - 18 noun cases
  - `házaitokban [in your (plural) houses]`
  - `láttunk [we saw (indefinite)]`
- Can neural networks perform morphological analysis?

Morphological analysis

- Finite-state transducers: finite-state machine with two memory tapes
- Lower tape: surface form of word
- Upper tape: morphological analysis of the word
- FSTs are reversible
- Large inventory of rules written by computational linguists
- Character/morpheme level input and output tapes

Morphological analysis

a	p	p	l	e	[/N]	s	[/Pl]

Synthesis ↓ ↑ **Analysis**

a	p	p	l	e		s	

Surface form

Encoder-decoder networks

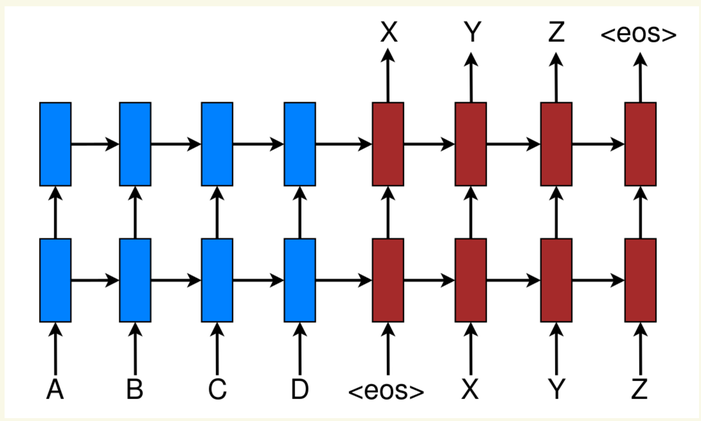

Figure: Standard sequence-to-sequence or encoder-decoder network. Source [2]

Luong's attention

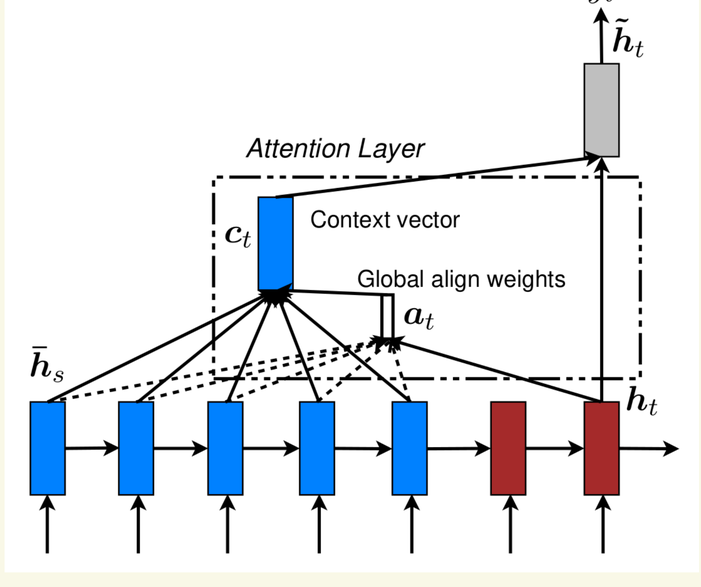

Figure: Luong's global attention model. Source [2]

Hard monotonic attention

- Aharoni and Goldberg, 2017 [1]
- Hard: attend a single source symbol at each time step
- Monotonic: limit the movement of the attention head to forward steps
  - STEP symbols in the target sequence move the head one source symbol forward
  - Requires STEP symbols in the target data
- Manual conversion from FST output:

FST output	a:a p:p l:l e:e :[/N] s:s :[Pl]
seq2seq source	a p p l e
seq2seq target	a **S** p **S** p **S** l **S** e [/N] **S** s [Pl]

- I implemented a batched version, massive spead-up.

Hungarian examples

Word	Analysis	English translation
jóváhagyásával	jóváhagyás[/N]a[Poss.3Sg]val[Ins]	with his/her approval
venne	vesz[/V]ne[Cond.NDef.3Sg]	he/she would buy sth [indefinite]
okosat	okos[/Adj]at[Acc]	(the) clever [accusative]
érezték	érez[/V]ték[Pst.Def.3Pl]	they felt [definite]

Experimental setup

- Dataset
  - Szeged Korpusz (gold standard morphological analysis) - 115k types
  - analyzed with emMorph [3]
  - disambiguated using the gold standard
- Architecture
  - Luong: 1 layer, 512 LSTM cells, 40/50 dimensional embeddings, no dropout
  - Hard monotonic: 2 layers, 512 LSTM cells, 32/32 dimensional embeddings, 0.5 dropout
  - Training: Adam, 0.0001 learning rate, maximum 100 epochs, early stopping
  - Model with lowest validation loss saved

Results

Table: Luong's attention

beam position			any correct
1	2	3	
0.7775	0.2273	0.0957	0.8603

Table: Hard monotonic attention

best model	majority vote
0.8831	**0.8992**

Conclusion and future work

- Make it reversible (as FSTs are)
- Other languages
- Do not rely on pre-existing STEP symbols

References

[1] Aharoni, Roee and Goldberg, Yoav,
Morphological inflection generation with hard monotonic attention.
ACL, 2017.

[2] Luong, Thang and Pham, Hieu and Manning, Christopher D.,
Effective Approaches to Attention-based Neural Machine Translation.
EMNLP, 2015.

[3] Attila Novák and Borbála Siklósi and Csaba Oravecz,
A New Integrated Open-source Morphological Analyzer for Hungarian.
LREC, 2016.

Judit Ács
judit@sch.bme.hu
Human Language Technologies Group, Institute for Computer Science and Control, Hungarian Academy of Sciences