Neural networks for transducing between DMRS graphs in English and Japanese

Ryan Pham
Michael Goodman
Jan Buys

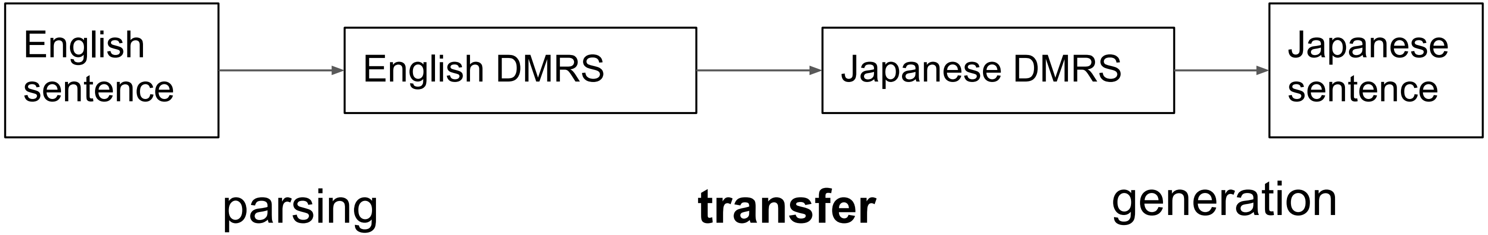
PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Towards Neural Machine Translation with MRS

- NMT is state-of-the-art for machine translation (for most language pairs)
- But using explicit syntax and/or semantics might still improve performance

Towards Neural Machine Translation with MRS

Semantic machine translation:



Semantic transfer

- Formulate as a graph transduction problem, using DMRS as representation
- Language pair where broad-coverage grammars are available for both languages: English (ERG) and Japanese (JACY)
- Use parsed parallel corpora as training data
- Evaluation metric: Smatch score (F1 graph overlap score)

Semantic transfer

- Use neural networks for graph transduction
- Formulate as a sequence-to-sequence problem: linearize graphs
- Train neural encoder-decoder model
- Does not require explicit alignments between the graphs

Graph linearization

Original

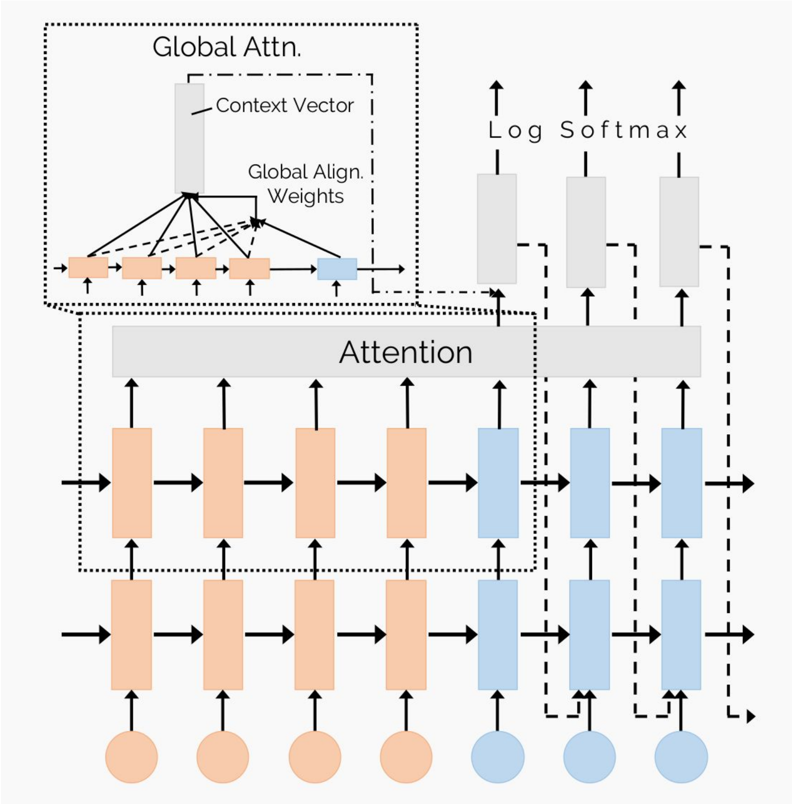
```
(10002 / _be_v_id
    :lnk "<4:6>"
    :sf prop
    :tense pres
    :mood indicative
    :prog -
    :perf -
    :cvarsort e
    :ARG1-NEQ (10000 / generic_entity
        :lnk "<0:4>"
        :pers 3
        :num sg
        :gend n
        :cvarsort x
        :RSTR-H-of (10001 / _that_q_dem
            :lnk "<0:4>"))
    :ARG2-NEQ (10005 / _idea_n_of
        :lnk "<14:19>"
        :pers 3
        :num sg
        :ind +
        :cvarsort x
        :RSTR-H-of (10003 / _a_q
            :lnk "<7:8>")
        :ARG1-EQ-of (10004 / _good_a_at-for-of
            :lnk "<9:13>"
            :sf prop
            :tense untensed
            :mood indicative
            :prog -
            :perf -
            :cvarsort e)))
```

That's a good idea!

Simplified

```
( _be_v_id
    :sf=prop:tense=...=-:cvarsort e
    :ARG1-NEQ( generic_entity
        :pers=3:num=...=n:cvarsort x
        :RSTR-H-of( _that_q_dem ) )
    :ARG2-NEQ( _idea_n_of
        :pers=3:num=...=+:cvarsort x
        :RSTR-H-of( _a_q )
        :ARG1-EQ-of( _good_a_at-for-of
            :sf=prop:tense=...=-:cvarsort e)))
```

Neural network architecture



- Encoder-decoder RNN with Global (Luong) attention and input feeding

Datasets

- Parse bitext with grammars, only use sentence pairs when both are parsable

Tanaka	Kyoto	Japanese WordNet
• Not *perfect* translations • ~ 124,000 pairs • Split in 2 sets	• *Exact* translations of Wikipedia articles • ~131,000 pairs	• Corpus modeled after WordNet • ~114,000 pairs

Results

Dataset	Variation	SMATCH	B.O.W.
Tanaka'	--	0.61	0.61
	- Features	0.63	0.62
	+ Coverage	0.61	0.61
	Transformer	0.59	0.59
JPN WN	--	0.57	0.54
All – Tanaka'	--	0.53	0.52
All	--	0.57	0.56

Results

Dataset	Model	SMATCH	Abstract Pred.		Surface Pred.	
			Precision	Recall	Precision	Recall
Tanaka’	–	0.61	0.75	0.70	0.55	0.48
	-Features	0.63	0.73	0.73	0.56	0.52
	+Coverage	0.61	0.75	0.70	0.56	0.49
	Transformer	0.59	0.77	0.65	0.56	0.46
JPN WN	–	0.57	0.83	0.61	0.52	0.39
All - Tanaka’	–	0.53	0.78	0.55	0.49	0.30
All	–	0.57	0.76	0.61	0.52	0.38

Example Prediction

Expected	Prediction
(_motsu_v	(_motsu_v
	:ARG1-NEQ(pron
	:RSTR-H-of(def_q)
	:ARG2-EQ-of(_wa_d))
:ARG2-NEQ(_hon_n	:ARG2-NEQ(_hon_n
:RSTR-H-of(_kono_q)	:RSTR-H-of(_kono_q)
:ARG1-EQ-of(_atarashii_a)))	:ARG1-EQ-of(_atarashii_a)))

Common Predicate Errors			
Abstract		**Surface**	
def_q	compound	_no_p	_sono_q
udef_q	named	_wa_d	_exist_v
nominalization	pron	_ni_p	_koto_n_nom

Error analysis

- Sequential model has trouble dealing with long-distance relations between elements in the graph
- Abstract predicates and versatile surface predicates are often mispredicted
- Training data size limited by coverage of the Japanese grammar

Future work

- Model architectures that are more suitable for graphs
  - TreeLSTM or graph convolutional encoders
  - Stack-based decoders
  - Parent feeding for dealing with long-distance dependencies
- Neural network approach requires more training data
  - Pre-trained monolingual models might also help
- Evaluation: Need upper bounds on performance, gold-annotated test data, methods to estimate training data quality

Future work

- Evaluation: Need upper bounds on performance, gold-annotated test data, methods to estimate training data quality
- Towards full translation systems: Need annotated data in (more) languages to train parsers and generators
- In low-resource settings, how can we make use of grammars with limited coverage?