Syntax-based Language Modelling with Recurrent Neural Network Grammars

Carnegie Mellon University

Austin Matthews, Alon Lavie,
Graham Neubig, Chris Dyer
austinma@cs.cmu.edu

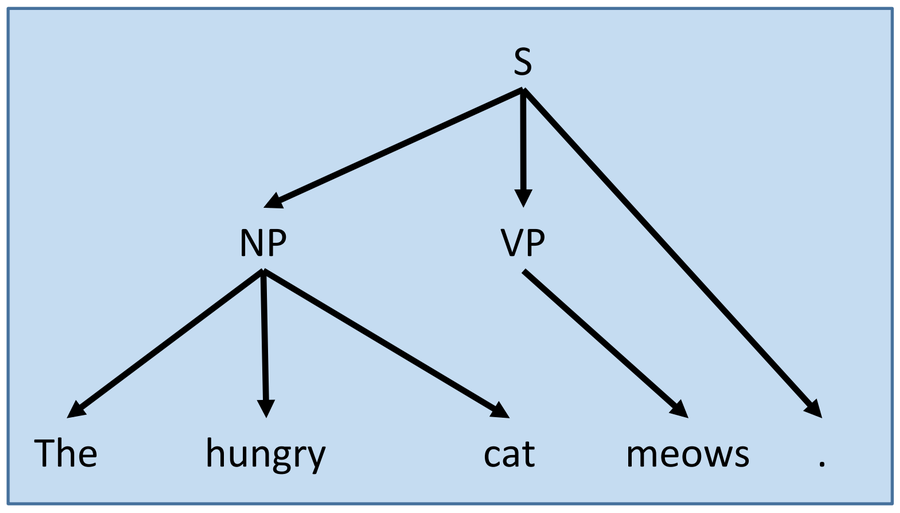
Language Technologies Institute

Recurrent Neural Network Grammars

Operation Sequence Model of Trees

Choose next action conditioned on stack, generated terminals, and action history.

Terminals	Stack	Action
		NT(S)
	(S	NT(NP)
	(S (NP	GEN(The)
The	(S (NP The	GEN(hungry)
The hungry	(S (NP The hungry	GEN(cat)
The hungry cat	(S (NP The hungry cat	REDUCE
The hungry cat	(S (NP The hungry cat)	NT(VP)
The hungry cat	(S (NP The hungry cat) (VP	GEN(meows)
The hungry cat meows	(S (NP The hungry cat) (VP meows	REDUCE
The hungry cat meows	(S (NP The hungry cat) (VP meows)	GEN(.)
The hungry cat meows .	(S (NP The hungry cat) (VP meows) .	REDUCE
The hungry cat meows .	(S (NP The hungry cat) (VP meows) .)	



- Simulates tree generation in a top-down, left-to-right fashion
- Maintains an explicit stack of (partial) constituents
- Learns a composition function that composes all elements on the stack up to the most recent open parenthesis into one vector representation

Marginalization

Importance sampling

RNNGs give us a model of p(tree, sentence), but often we want to compute p(sentence), marginalizing over trees.

$$p(s) = \sum_{t \in \mathcal{T}} p(s,t)$$ Definition of marginal probability

$$= \sum_{t \in \mathcal{T}} p(s,t) \cdot \frac{q(t|s)}{q(t|s)}$$ Multiply by 1

$$= \sum_{t \in \mathcal{T}} q(t|s) \cdot \frac{p(s,t)}{q(t|s)}$$ Rearrange

$$= \mathbb{E}_{t \sim q(t|s)} \frac{p(s,t)}{q(t|s)}$$ Definition of expectation

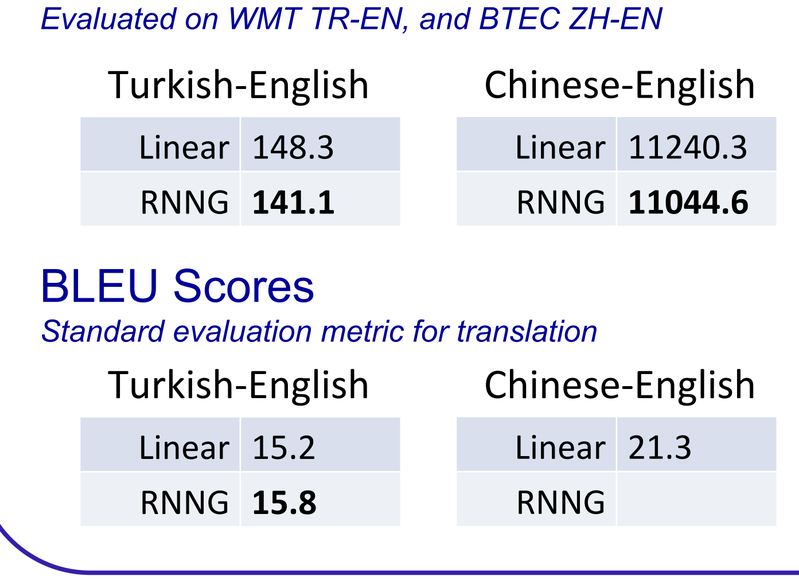
$$\approx \sum_{i=1}^{N} \frac{p(s,t_i)}{q(t_i|s)}$$ Monte Carlo approximation

- Use a discriminative parser (e.g. Berkeley, Stanford) to sample trees *t* and their probabilities *q(t | s)*.
- Allows RNNGs to be compared to or used in place of non-syntactic LMs

Source Conditioning

RNNG Translation Models

Attention, which is now standard in NMT, can be used to turn an RNNG LM into a translation model.



- Use the current parser state to predict weights for each source word
- Average word vectors according to the chosen weights
- Predict next action conditioned on the current parser state and context vector

Language Modeling Results

Marginalized Perplexity on Tree Banks

Evaluated perplexity on the Penn Tree Bank and Chinese Tree Bank using importance sampling marginalization.

Translation Results

Marginalized Perplexity on Translation

Evaluated on WMT TR-EN, and BTEC ZH-EN

Turkish-English	
Linear	148.3
RNNG	**141.1**

Chinese-English	
Linear	11240.3
RNNG	**11044.6**

BLEU Scores

Standard evaluation metric for translation

Turkish-English	
Linear	15.2
RNNG	**15.8**

Chinese-English	
Linear	21.3
RNNG	