BiDeT
Bidirectionally Decoding Translator

Ganesh Murugappan and Tejas Lokeshrao

Motivation

- Encoders are generally bidirectional. Why not decoders?
- Decoders can get stuck if they start poorly
- Provide context from both sides

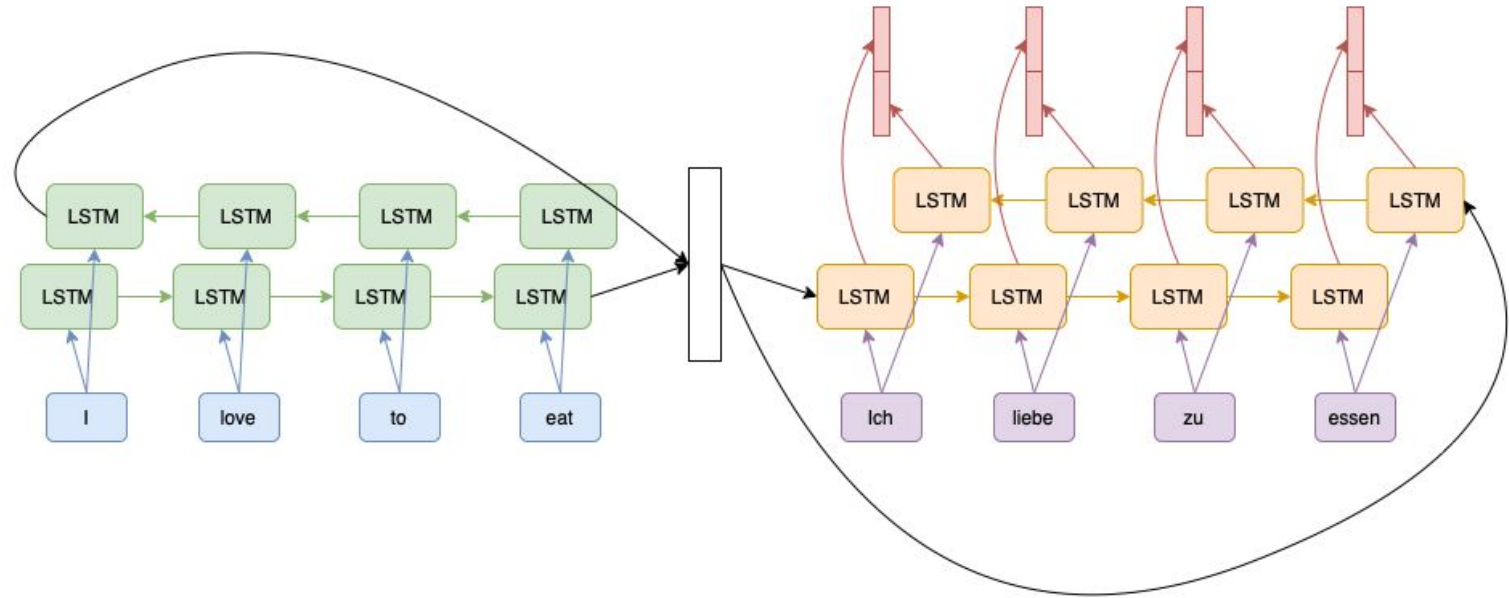BiDeT: Bidirectionally Decoding Translator

Ideation

- Wanted to implement bidirectional transformer decoder
- Instead, use simpler, autoregressive base: LSTM
- What if left-to-right and right-to-left outputs vary in length?

BiDeT: Bidirectionally Decoding Translator

Bidirectional Decoding

BiDeT: Bidirectionally Decoding Translator

Data: WMT14 English-to-German

- 4.5 million pairs of English and German sentences
- Validation and test sets of 3000 pairs each
- Words that occur only once are removed from vocabulary

BiDeT: Bidirectionally Decoding Translator

Experimental Setup

- Bidirectional LSTM encoder
- Bidirectional LSTM decoder!
- 100,000 samples used in training without repetition
- Backpropagation on cross entropy loss

BiDeT: Bidirectionally Decoding Translator

Results

Model size	# parameters
UniDeT	30M
BiDeT	43M

Cross entropy loss	Train	Validation	Test
UniDeT	1.369	1.543	1.656
BiDeT	1.085	1.191	1.301

BiDeT: Bidirectionally Decoding Translator

Future Work

- Implement bidirectional decoder in transformer
- Get BLEU scores to compare model performance in inference

BiDeT: Bidirectionally Decoding Translator