

Neural Machine Translation (NMT)

Presenter: Noriyuki Kojima

Date: 2017/10/3

Machine Translation (MT)

“sub-field of computational linguistics that investigates ***the use of software*** to translate text or speech from one language to another.”

By wikipedia

What software is available for MT

- Simple Word Replacement
 - Apparent Low Accuracy
- Phrase-based Translation (Corpus Statistics)
 - N-Gram and EM-algorithm, it is not terrible ...
- Neural Machine Translation
 - Use of Neural Network, best performance (En-Fr)

VS

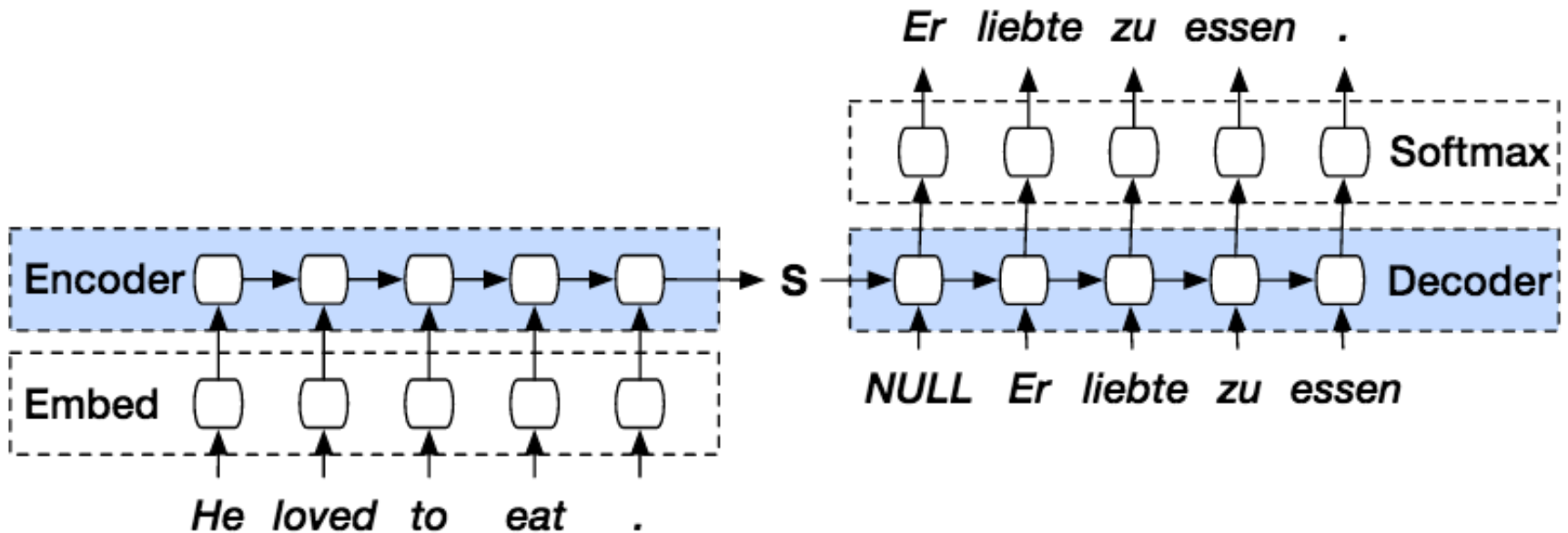
- Human Translation

Steps to understand NMT

1. Encoder-Decoder
2. Attention-Based Encoder-Decoder
3. Bi-directional Encoder-Decoder
4. Depth
5. Testing (Beam Search)
6. Cool New Models of NMT

1. Encoder-Decoder

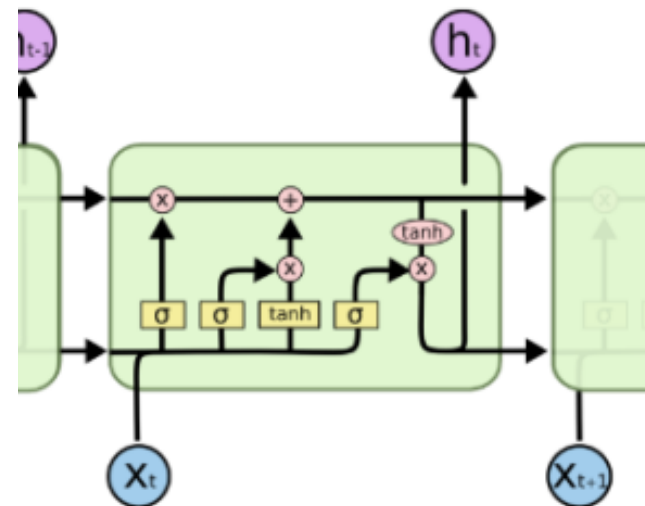
- Encoding?/ Decoding?
- Hidden state



1.Encoder-Decoder

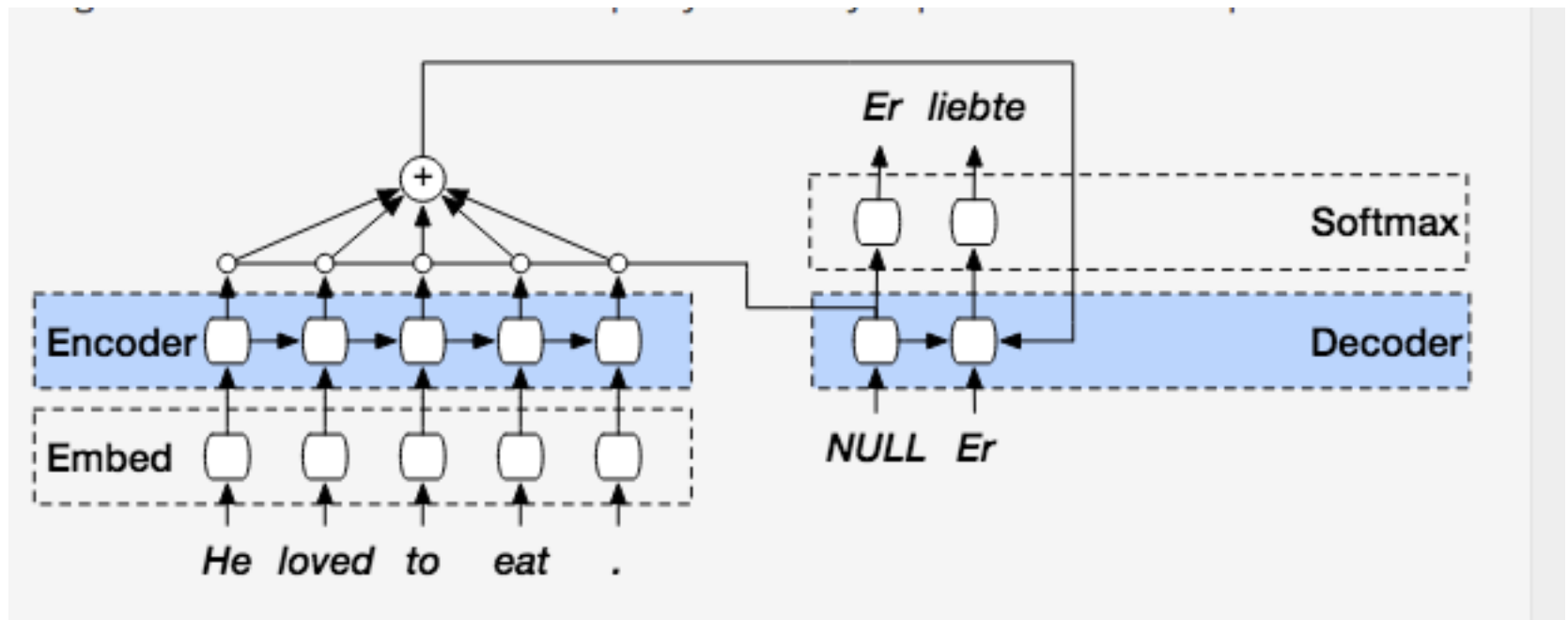
- Recurrent Neural Network
- Recurrent Cell Unit (Mi-LSTM/LSTM/GRU)
 - Capture the dependency of a sequence as Memory
 - Three components
 - Memory (hidden state) + Input \rightarrow New Memory
 - Input Gate, Output Gate, Forget Gate
 - Parameter Sharing

Good Resource: <https://theneuralperspective.com/2016/11/17/recurrent-neural-network-rnn-part-4-custom-cells/>



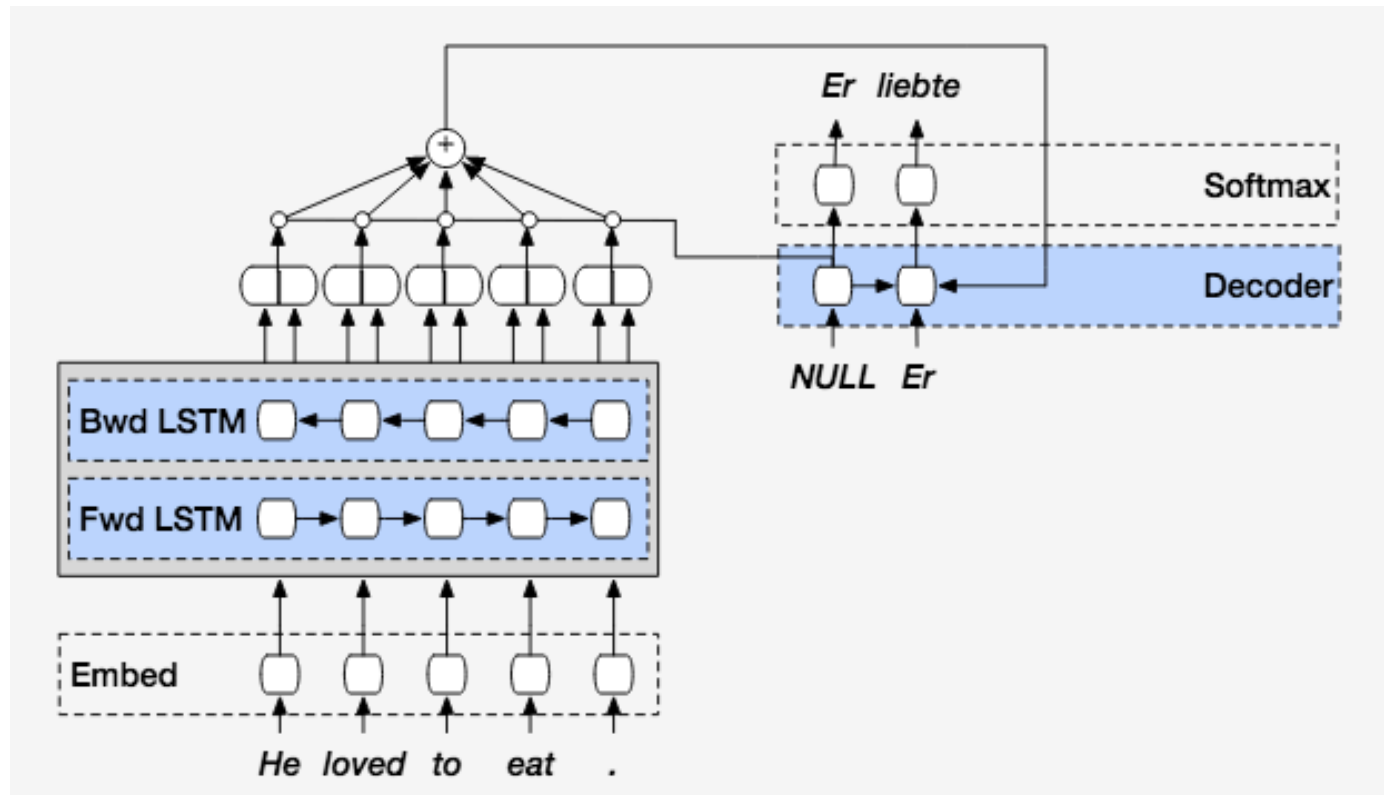
2. Attention-Based Encoder-Decoder

- Problem of RNN: Vanishing Gradients
- Solution: Attention!!!



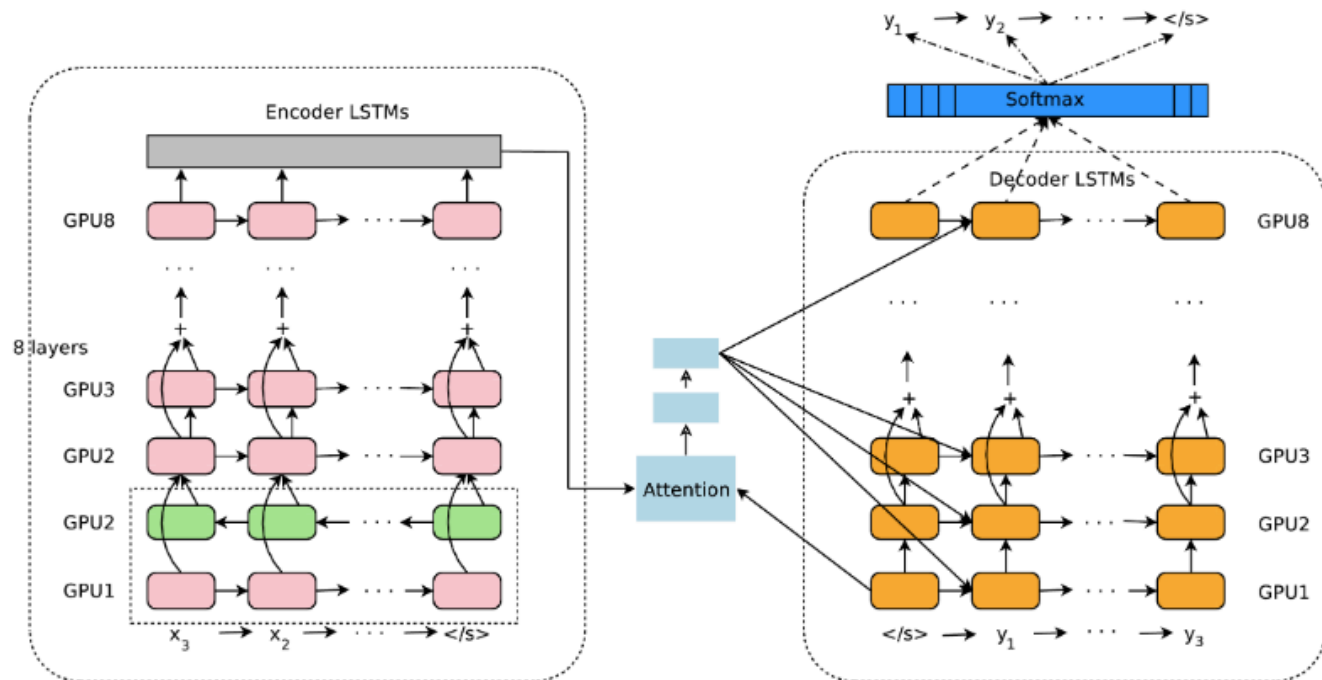
3. Bi-directional Encoder-Decoder

- Capture the full context
 - “I have a car” or Arabic



4.Depth

- Depth sometimes help..... (but not always)
- GNMT



5. Beam Search (Testing)

Recall output of Decoder is probability of word.....

Need to calculate probability of sequence

→ Beam Search!!

5. Beam Search (Testing)

	/s							
Probability	1							
	I	He	She					
Probability	0.4	0.1	0.5					
	Have	Run	Have			Has	Run	Have
Probability	0.1	0.7	0.2		Probability	0.5	0.45	0.05

$$P(\text{Top1}) = P(/s \mid \text{Run}) = 1 * 0.4 * 0.7 = 0.28$$

$$P(\text{Top2}) = P(/s \text{ She has}) = 1 * 0.5 * 0.5 = 0.25$$

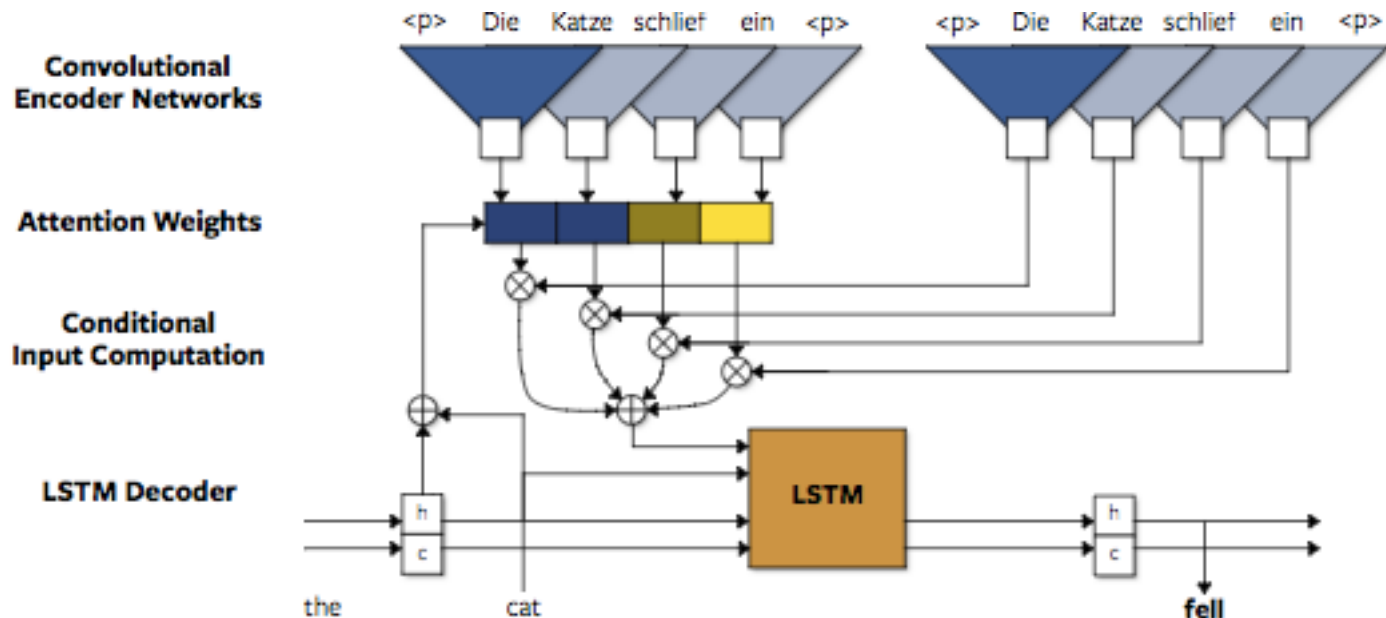
$$P(\text{Top3}) = P(/s \text{ She Run}) = 1 * 0.5 * 0.45 = 0.225$$

.

.

6. Cool New Models

- FAIR CNN encoder-(decoder)



6. Cool New NMTs

- Google's Multilingual NMT (Johnson et al., 2017)
 - Many to one, one to many, many to many
- Adversarial NMT (Wu et al., 2017)
 - GANs for NMT
- Learning to Translate in Real-time with NMT (Gu et al., 2017)
 - Reinforcement Learning for NMT
- Context aware, low-resource, BLUE, Ensemble and etc...

Thank you for listening!!!