Neural Machine Translation by Jointly Learning to Align and Translate

Name • Maab Elrashid Ahmed Mohamed

Outline

1. Core idea of the paper.
2. Novel model architecture.
3. Implementation approach.
4. Personal experiment results.
5. Difference comparison with the original results.
6. Challenges.

Idea

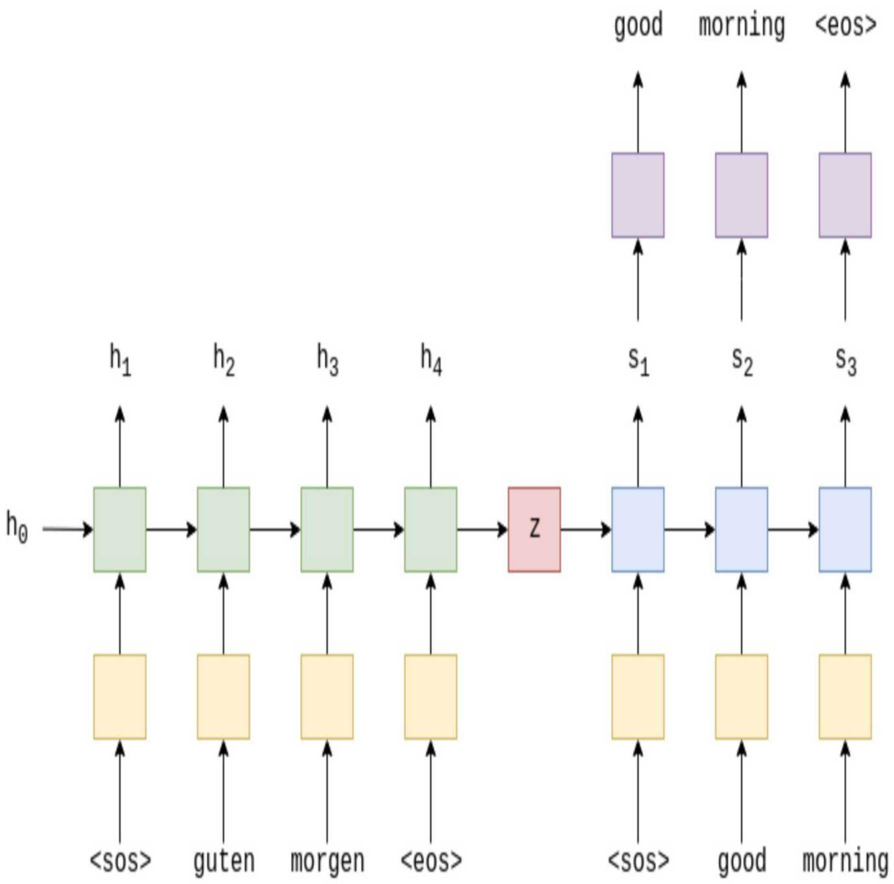
- Basic RNN encoder-decoder model and its limitation.
- The proposed model; RNNsearch, and it addresses the issue by 2 ways:

a) using bidirectional RNN for the input.

b)introducing an Alignment model (attention mechanism).

Architecture

RNN enc-dec



good
morning
<eos>
h_1
h_2
h_3
h_4
s_1
s_2
s_3
h_0
z
<sos>
guten
morgen
<eos>
<sos>
good
morning

RNN enc-dec

Encoder

$$h_t = f\left(x_t, h_{t-1}\right)$$

$$c = q\left(\{h_1, \cdots, h_{T_x}\}\right),$$

Decoder

$$p(\mathbf{y}) = \prod_{t=1}^{T} p(y_t \mid \{y_1, \cdots, y_{t-1}\}, c),$$

RNN search

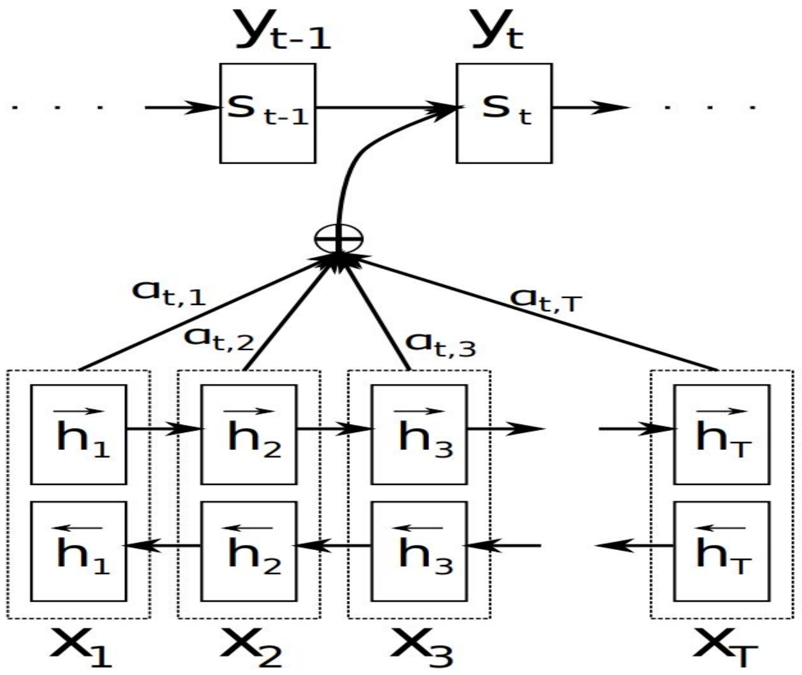


Figure 1: The graphical illustration of the proposed model trying to generate the t-th target word y_t given a source sentence $(x_1, x_2, \ldots, x_T)$.

Encoder: Bidirectional RNN for Annotating Sequences

$$h_j = \left[\overrightarrow{h}_j^\top ; \overleftarrow{h}_j^\top \right]^\top$$

Decoder

Alignment model

$$e_{ij} = a(s_{i-1}, h_j)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

Decoder LSTM

$$s_i = f(s_{i-1}, y_{i-1}, c_i)$$

$$p(y_i | y_1, \ldots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i)$$

Experiment Settings

Paper Experiment

Task

English to French translation.

Data

The bilingual, parallel corpora provided by ACL WMT '14. (Europarl (61M words), news commentary (5.5M), UN (421M) and two crawled corpora of 90M and 272.5M words respectively, totaling 850M words).

Code

The implementation of the paper available here.

Personal Experiment

Task

English-German translation.

Data

Using torchtext dataset for machine translation (Multi30k). And use spacy for the tokenization of the data.

Implementation:

Available here.

Results

Paper results

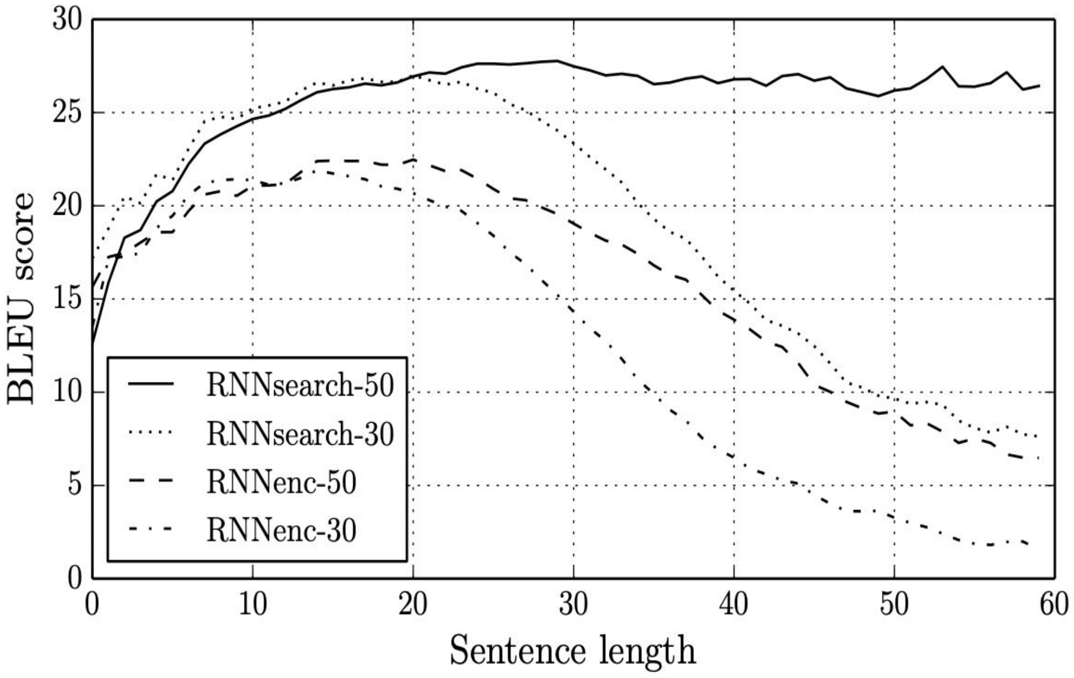
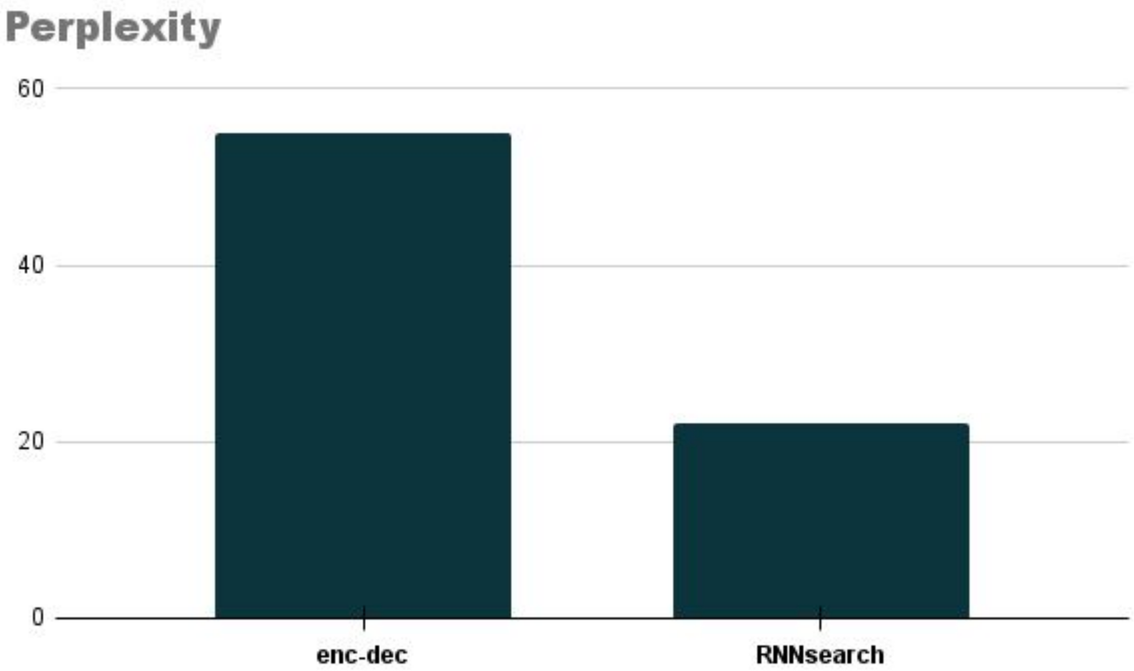


Figure 2: The BLEU scores of the generated translations on the test set with respect to the lengths of the sentences. The results are on the full test set which includes sentences having unknown words to the models.

Personal results



The lower the perplexity, the better the model is.

The Left Challenge

Better handle the unknown or rare words, and to improve the translation.

Thank you..