

# Global Encoding for Abstractive Summarization

Kefan Xu

School of Electronics Engineering and Computer Science

Peking University

Advisor:  
Baobao Chang(PKU)

# Outline

- Motivaton
- CNN Gate Unit
- Global Attention Unit
- Final Result

# I. Motivation — Difficulty

- **Disadvantages:**

(1) repetition

(2) semantic irrelevance

- **Cause:**

Focusing on the local information too much

- **Our solution:**

Proposed an seq-to-seq model with dealing with global information

---

**Text:** the mainstream fatah movement on monday officially chose mahmoud abbas, chairman of the palestine liberation organization (plo), as its candidate to run for the presidential election due on jan. #, ####, the official wafa news agency reported.

---

**seq2seq:** fatah officially officially elects abbas as candidate for candidate .

---

**Gold:** fatah officially elects abbas as candidate for presidential election

---

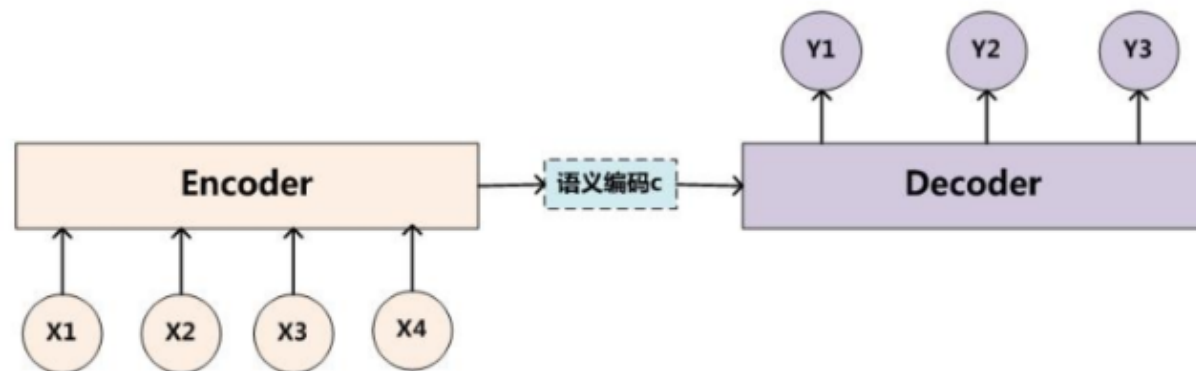
seq2seq

Global info

Solve local-focus problem

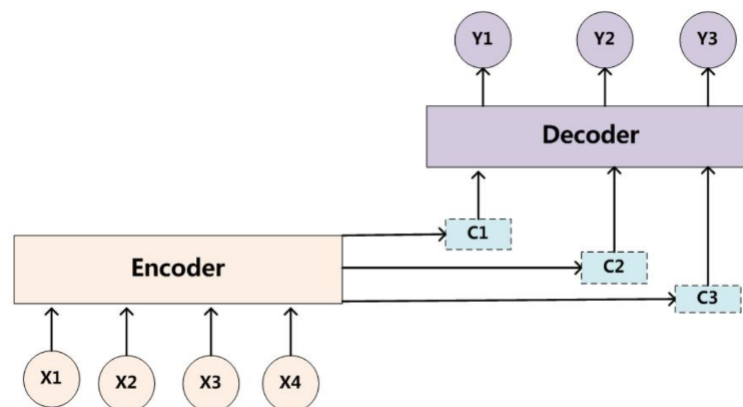
# I. Motivation — original model

- The encoder-decoder model is widely used to finish the machine abstractive summarizing or translating tasks.
- **Encoder :**
  - (1) Extract the meaning of the input
  - (2) Generate a context vector.
- **Decoder:**
  - Generate the summarization or the translation based on the the context information produced by the Encoder.



## II. Global Attention model

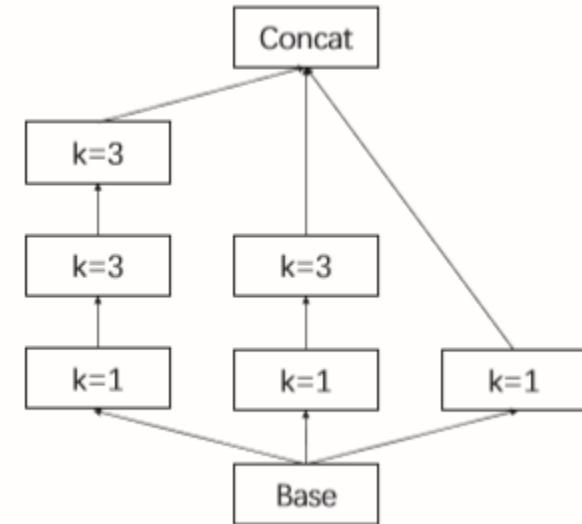
- **One of the serious problem of the encoder-decoder:**
- While generating the different outputs, the decoder using the same context
- the model may likely focus on similar aspects of the context information and produce the outputs similar with others.
- **Solution:**
- Attention model means provide different context information while dealing with various outputs.
- Help decoder focus on the more related context information.



$$P_{vocab} = softmax(g([c_t; s_t]))$$
$$s_t = LSTM(y_{t-1}, s_{t-1}, C_{t-1})$$
$$c_t = \sum_{i=1}^n \alpha_{t,i} h_i$$
$$\alpha_{t,i} = \frac{exp(e_{t,i})}{\sum_{j=1}^n exp(e_{t,j})}$$
$$e_{t,i} = s_{t-1}^\top W_a h_i$$

# III. CNN Gate Unit

- **Attention model** alleviate the influence of the repetition
- **But** it don't help address the problem of the semantic irrelevance .
- **Solution:**
- Provided a gate unit above the output of the attention model.
- Used serval CNN with different size of kernels and concat the final outputs
- Help the model learn about n-gram information, and perform the outputs with reasonable structure.



# IV. Result

Model	R-1	R-2	R-L
RNN	21.5	8.9	18.6
RNN-context	29.9	17.4	27.2
CopyNet	34.4	21.6	31.3
SRB	33.3	20.0	30.1
DRGD	37.0	24.2	34.2
seq2seq (Our impl.)	33.8	23.1	32.5
<b>+CGU</b>	<b>39.4</b>	<b>26.9</b>	<b>36.5</b>

Table 2: **F-Score of ROUGE on LCSTS.**

Model	R-1	R-2	R-L
ABS	29.6	11.3	26.4
ABS+	29.8	11.9	27.0
Feats	32.7	15.6	30.6
RAS-LSTM	32.6	14.7	30.0
RAS-Elman	33.8	16.0	31.2
SEASS	36.2	17.5	33.6
DRGD	<b>36.3</b>	17.6	33.6
seq2seq (Our impl.)	33.6	16.3	31.3
<b>+CGU</b>	<b>36.3</b>	<b>18.0</b>	<b>33.8</b>

Table 3: **F-Score of ROUGE on Gigaword.**