



Sequential Copying Networks

Qingyu Zhou^{1*}, Nan Yang², Furu Wei², Ming Zhou²

¹Harbin Institute of Technology, Harbin, China ²Microsoft Research, Beijing, China

qyzhgm@gmail.com {nanya, fuwei, mingzhou}@microsoft.com

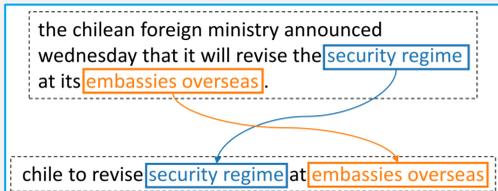
* Work done during an internship at Microsoft Research.

Why Sequential Copying?

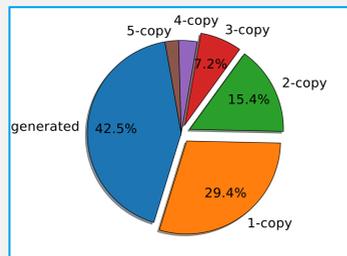
Abstractive summarization as an example:

Sequential Copying happens when doing abstractive summarization:

- Copying a span from input sentence



An example of sequential copying in abstractive sentence summarization task.



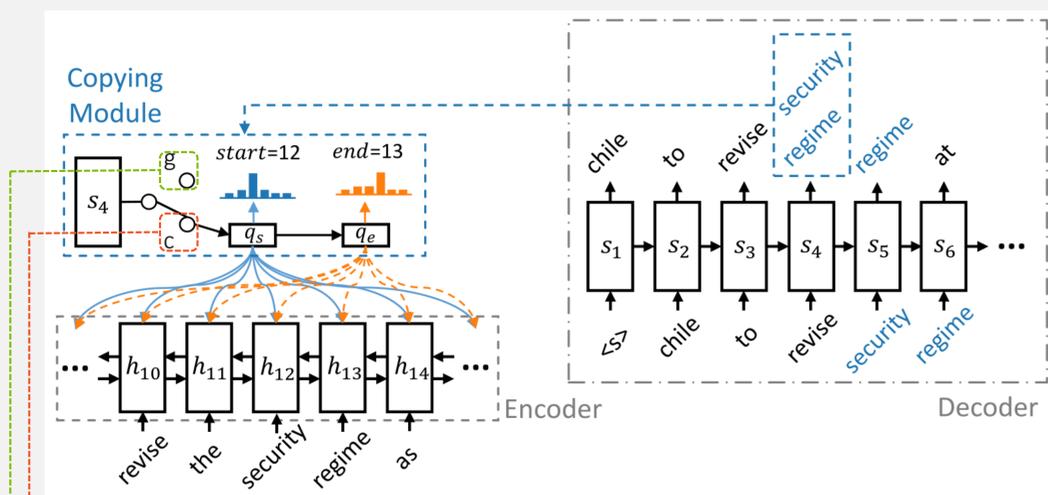
Percentage of generated and copied words in sentence summarization training data.

Sequential Copying is ESSENTIAL for such tasks and datasets

Sequential Copying Networks (SeqCopyNet)

Problem: "Single word copying" paradigm may introduce errors due to separate copying decisions when copying multi-words

Solution: We copy them once and for all



Sequential Copying Module:

- Copy Switch Gate
- Span Pointer Network
- Copy State Transducer

$$\text{Decoder memory: } m_t = \begin{bmatrix} y_{t-1} \\ s_t \\ c_t \end{bmatrix}$$

Copy Switch Network:

$$G(x) = \sigma(W_2(\tanh(W_1x + b_1)) + b_2)$$

Copy Mode: $p_c = G(m_t)$ → Copy a span from input

Generate Mode: $p_g = 1 - p_c$ → Generate next word from target vocabulary

Two-step Span Copying:

$$q_s = \tanh(W_s m_t + b)$$

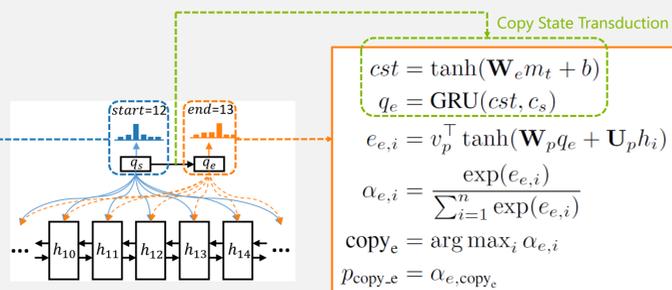
$$e_{s,i} = v_p^T \tanh(W_p q_s + U_p h_i)$$

$$\alpha_{s,i} = \frac{\exp(e_{s,i})}{\sum_{i=1}^n \exp(e_{s,i})}$$

$$\text{copy}_s = \arg \max_i \alpha_{s,i}$$

$$p_{\text{copy}_s} = \alpha_{s, \text{copy}_s}$$

$$c_s = \sum_{i=1}^n \alpha_{s,i} h_i$$



$$cst = \tanh(W_e m_t + b)$$

$$q_e = \text{GRU}(cst, c_s)$$

$$e_{e,i} = v_p^T \tanh(W_p q_e + U_p h_i)$$

$$\alpha_{e,i} = \frac{\exp(e_{e,i})}{\sum_{i=1}^n \exp(e_{e,i})}$$

$$\text{copy}_e = \arg \max_i \alpha_{e,i}$$

$$p_{\text{copy}_e} = \alpha_{e, \text{copy}_e}$$

Experiments

Task 1: Question Generation give a sentence and its desired answer

Dataset: Question Generation dataset based on SQuAD.

Evaluation Metric: BLEU-4

Model	Dev set	Test set
PCFG-Trans [‡]	9.28	9.31
s2s+att [‡]	3.01	3.06
NQG [‡]	10.06	10.13
NQG+ [‡] (single copy)	12.30	12.18
SeqCopyNet	13.13	13.02

Task 2: Abstractive Sentence Summarization

Dataset: English Gigaword: Rush, Chopra, and Weston (2015)*, Zhou et al. (2017b) and our internal test sets

Evaluation Metric: ROUGE F1

Models	Test set in Zhou et al. (2017b)			Our internal test set		
	RG-1	RG-2	RG-L	RG-1	RG-2	RG-L
ABS [‡]	37.41	15.87	34.70	-	-	-
s2s+att (greedy)	46.21	24.02	43.30	45.46	22.83	42.66
s2s+att (beam)	47.08	25.11	43.81	46.54	24.18	43.55
NMT + UNK_PS (greedy)	45.64	23.38	42.67	45.21	23.01	42.38
NMT + UNK_PS (beam)	47.05	24.82	43.87	46.52	24.41	43.58
SEASS (greedy) [‡]	45.27	22.88	42.20	-	-	-
SEASS (beam) [‡]	46.86	24.58	43.53	-	-	-
SeqCopyNet (greedy)	46.51	24.14	43.20	46.08	23.99	43.26
SeqCopyNet (beam)	47.27	25.07	44.00	47.13	24.93	44.06

* Test sets of Rush, Chopra, and Weston (2015) and Zhou et al. (2017b) are similar, so we only show results on Zhou et al. (2017b) in this poster.

- SeqCopyNet outperforms "single word copy" baseline (NMT + UNK_PS)
- Rare words have already been replaced by <unk> in test sets of Rush, Chopra, and Weston (2015) and Zhou et al. (2017b), therefore both "single word copy" and SeqCopyNet performs comparably to the s2s_att baseline.
- We construct a new test set, in which the sentence-summary pairs are remained untouched (released on <https://res.qyzhou.me/>)

Example Output

Input:	david ortiz homered and scored three times , including the go-ahead run in the eighth inning , as the boston <u>red sox</u> beat the toronto <u>blue jays 10-9</u> in the american league on tuesday .
Reference:	david ortiz helps red sox beat blue jays 10-9
SingleCopy:	ortiz homers as red sox beat blue jays
SeqCopyNet:	[red sox] beat [blue jays 10-9]
Input:	guyana 's president <u>cheddi jagan</u> , a long-time marxist turned free - marketeer , died here early thursday , an embassy spokeswoman said . he was 78 .
Reference:	guyana 's president cheddi jagan marxist turned marketeer dies at 78
SingleCopy:	guyana 's president jagan dies at 78
SeqCopyNet:	[guyana 's president cheddi jagan] dies at 78
Input:	china topped myanmar 's marine <u>product exporting countries annually</u> in the past decade among over 40 's , the local voice weekly quoted the marine products producers and exporters association as reporting sunday .
Reference:	china tops myanmar's marine product exporting countries in past
SingleCopy:	china tops myanmar 's marine product export
SeqCopyNet:	china tops myanmar 's marine [product exporting countries annually]

SeqCopyNet can copy long spans from input sentence

SeqCopyNet is good at detecting boundaries:

- named entities
- noun phrases

After copying a long span, the decoder can still generate well

Input:	<u>peyton manning</u> became the first quarterback ever to lead two different teams to multiple super bowls .
Reference:	how many teams has manning played for that reached the super bowl , while he was on their team ?
SeqCopyNet:	how many teams did [peyton manning] lead ?
Input:	it is conjectured that a progressive <u>decline in hormone levels</u> with age is partially responsible for weakened <u>immune</u> responses in aging individuals .
Reference:	what is partially responsible for weakened immune response in older individuals ?
SeqCopyNet:	what is [responsible for weakened immune] responses in aging individuals ?
Input:	<u>the sarah jane adventures</u> , starring elisabeth sladen who reprised her role as investigative journalist sarah jane smith , was developed by cbbc ; a special aired on new year's day 2007 and a full series began on 24 september 2007 .
Reference:	when did the sarah jane series begin ?
SeqCopyNet:	on what date did [the sarah jane adventures] begin ?

Conclusion

- SeqCopyNet enables multi-word span copying, and can be integrated with seq2seq framework
- SeqCopyNet is good at detecting boundaries, such as named entity
- We release a new abstractive sentence summarization test set
- Future work: apply SeqCopyNet to other tasks such as dialogue generation