Get To The Point: Summarization with Pointer-Generator Networks

Abigail See, Peter J. Liu, Christopher D. Manning

Stanford University & Google Brain-ACL17

outline

- Background

- Motivation

- Model

- Comparison

- Experiments

- Results

Background

- Extractive methods: assemble summaries exclusively from passages

- Abstractive methods: generate novel words and phrases not featured in the source text

Background

- single-sentence summaries
  (e.g. headline generation tasks)
- multi-sentence summaries
  - higher levels of abstraction
  - consistency
  - avoiding repetition

Motivation

Two shortcomings(multi-sentence summaries):

- reproduce factual details inaccurately and inability to deal with out-of-vocabulary
- tend to repeat themselves

Two solution:

- copying words from the source text via pointing
- coverage vector

Model

Sequence-to-sequence attentional model

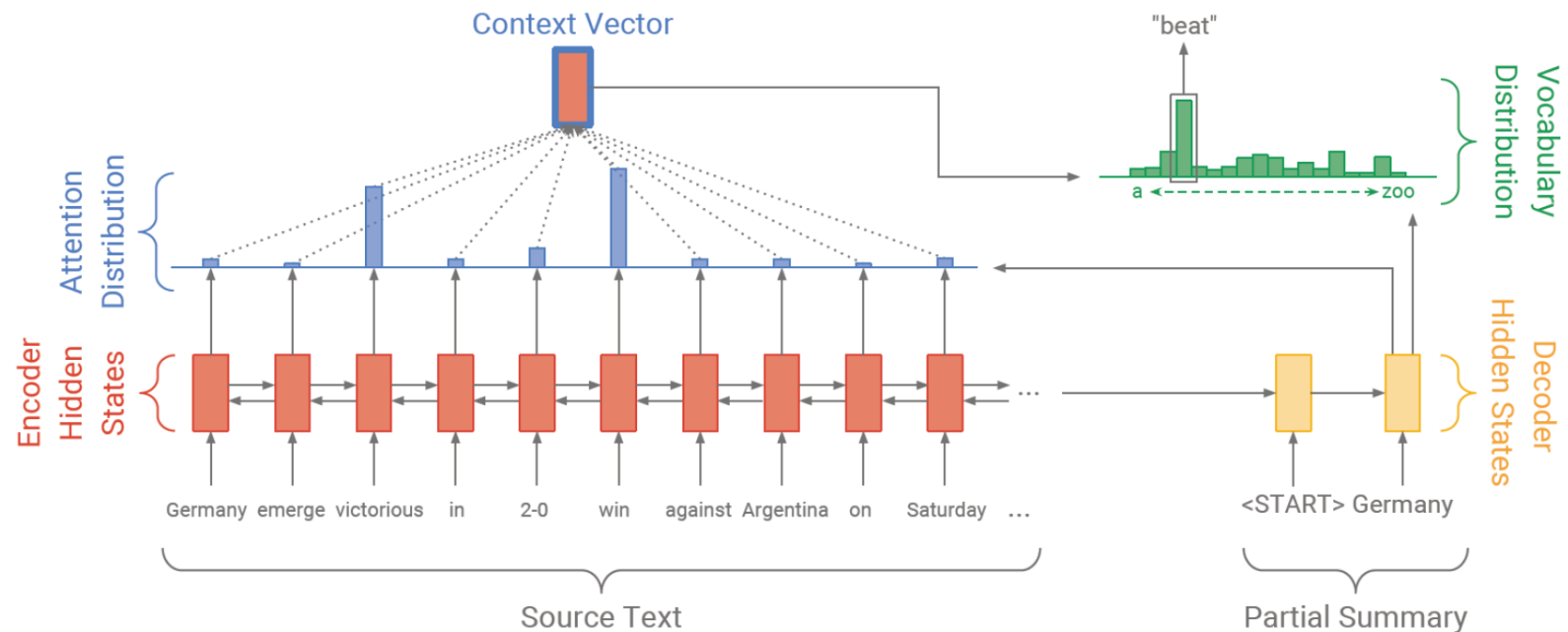Figure: Baseline sequence-to-sequence model with attention.

Seq2Seq + attentional model

$$e_i^t = v^T tanh(W_h h_i + W_s s_t + b_{attn})$$

$$a^t = softmax(e^t)$$

$$h_t^* = \sum a_i^t h_i$$

$$P_{vocab} = softmax(V^{'}(V[h_t^*, s_t] + b) + b^{'})$$

$$P(w) = P_{vocab}(w)$$

$$loss_t = -logP(w^*)$$

Seq2Seq + attentional model

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, *muhammadu buhari* told cnn's christiane amanpour that **he plans to aggressively fight corruption that has long plagued nigeria** and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, **he said his administration is confident it will be able to thwart criminals** and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

Baseline Seq2Seq + Attention: UNK UNK says his administration is confident it will be able to **destabilize nigeria's economy**. UNK says his administration is confident it will be able to thwart criminals and other **nigerians**. **he says the country has long nigeria and nigeria's economy.**
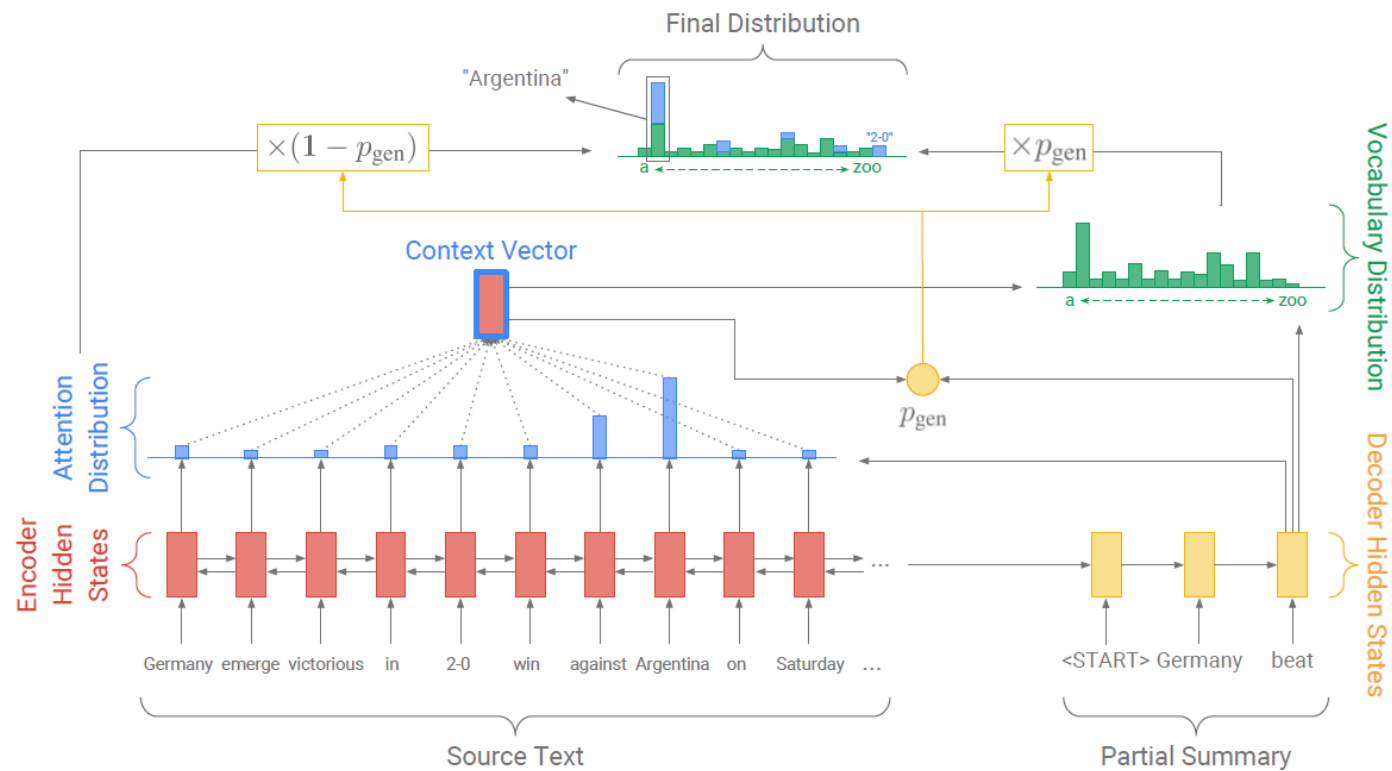
8

Pointer-generator network

Figure: Pointer-generator model.

Pointer-generator network

$$p_{gen} = \sigma(w_{h^*}^T h_t^* + w_{s_t}^T s_t + w_{x_t}^T x_t + b_{gen})$$

$$P(w) = p_{gen} P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i=w} a_i^t$$

generation probability: $p_{gen} \in [0, 1]$

Pointer-generator network

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, *muhammadu buhari* told cnn's christiane amanpour that **he plans to aggressively fight corruption that has long plagued nigeria** and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, **he said his administration is confident it will be able to thwart criminals** and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

Pointer-Gen: *muhammadu buhari* says he plans to aggressively fight corruption **in the northeast part of nigeria**. he says he'll "rapidly give attention" to curbing violence **in the northeast part of nigeria**. he says his administration is confident it will be able to thwart criminals.

Coverage mechanism

$$c_t = \sum_{t'=0}^{t-1} a_{t'}$$

$$e_i^t = v^T tanh(W_h h_i + W_s s_t + w_c c_i^t + b_{attn})$$

$$covloss_i = \sum_i min(a_i^t, c_i^t)$$

$$loss_t = -logP(w^*) + \lambda \sum_i min(a_i^t, c_i^t)$$

Pointer-Gen + Coverage

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, *muhammadu buhari* told cnn's christiane amanpour that **he plans to aggressively fight corruption that has long plagued nigeria** and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, **he said his administration is confident it will be able to thwart criminals** and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

Pointer-Gen + Coverage: *muhammadu buhari* says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

Pointer-Gen + Coverage

Coverage eliminates undesirable repetition
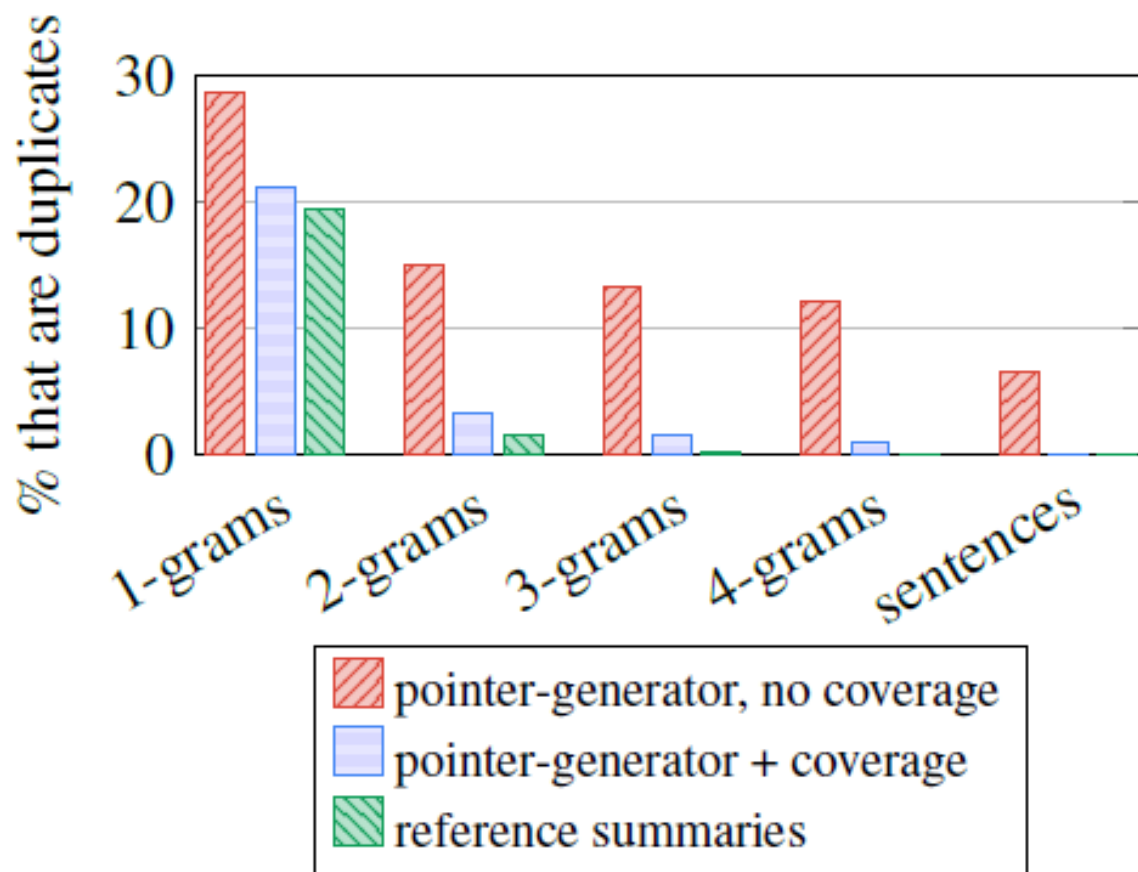
Comparison

CopyNet model V.S. Pointer-generator network

- explicit switch probability p_{gen} V.S shared softmax function

- copy distribution: recycle the attention distribution V.S two separate distributions

Forced-Attention model V.S. Pointer-generator network

- word appears multiple times: sum probability mass from all corresponding parts of the attention distribution

Comparison

Temporal attention V.S Coverage

- **Temporal attention**: each attention distribution is divided by the sum of the previous.
- early intervention method such as coverage is preferable to a post hoc method such as temporal attention.

Datasets

CNN/Daily Mail dataset

- only two published results on the full dataset

- articles: 781 tokens on average summaries: 3.75 sentences or 56 tokens on average

Metric

- ROUGE metric: F1 scores for ROUGE-1, ROUGE-2 and ROUGE-L

- METEOR metric: exact match mode, full mode

Experiments

- [**Setting**]
  - Adagrad initial accumulator: 0.1
  - learning rate: 0.15
  - truncate length: article 400 tokens, summary 100 tokens(train) 120(test)
  - batch size of 16
  - beam size 4 (test)
- [**Trips**]
  - start with highly-truncated sequences, then raise the maximum length once converged.
  - added the coverage mechanism with coverage loss weighted to $\lambda = 1$ further 3000 iterations

Results

- [**Baseline**]

  - lead-3 baseline: first three sentences of the article as a summary

  - only existing abstractive (Nallapati et al.,2016) and extractive (Nallapati et al., 2017) models on the full dataset.

    note: they generate anonymized summaries, this paper generate plain-text summaries

Observations

	ROUGE			METEOR	
	1	2	L	exact match	+ stem/syn/para
abstractive model (Nallapati et al., 2016)*	35.46	13.30	32.65	-	-
seq-to-seq + attn baseline (150k vocab)	30.49	11.17	28.08	11.65	12.86
seq-to-seq + attn baseline (50k vocab)	31.33	11.81	28.83	12.03	13.20
pointer-generator	36.44	15.66	33.42	15.35	16.65
pointer-generator + coverage	**39.53**	**17.28**	**36.38**	17.32	18.72
lead-3 baseline (ours)	40.34	17.70	36.57	20.48	22.21
lead-3 baseline (Nallapati et al., 2017)*	39.2	15.7	35.5	-	-
extractive model (Nallapati et al., 2017)*	39.6	16.2	35.3	-	-

- convincingly surpassing the best abstractive model
- does not quite surpass the ROUGE scores of the lead-3 baseline, nor the current best extractive model

Discussion

Comparison with extractive systems

- news articles tend to be structured with the most important information at the start

- the nature of the task and the ROUGE metric make extractive approaches difficult to beat

Discussion

How abstractive is our model?
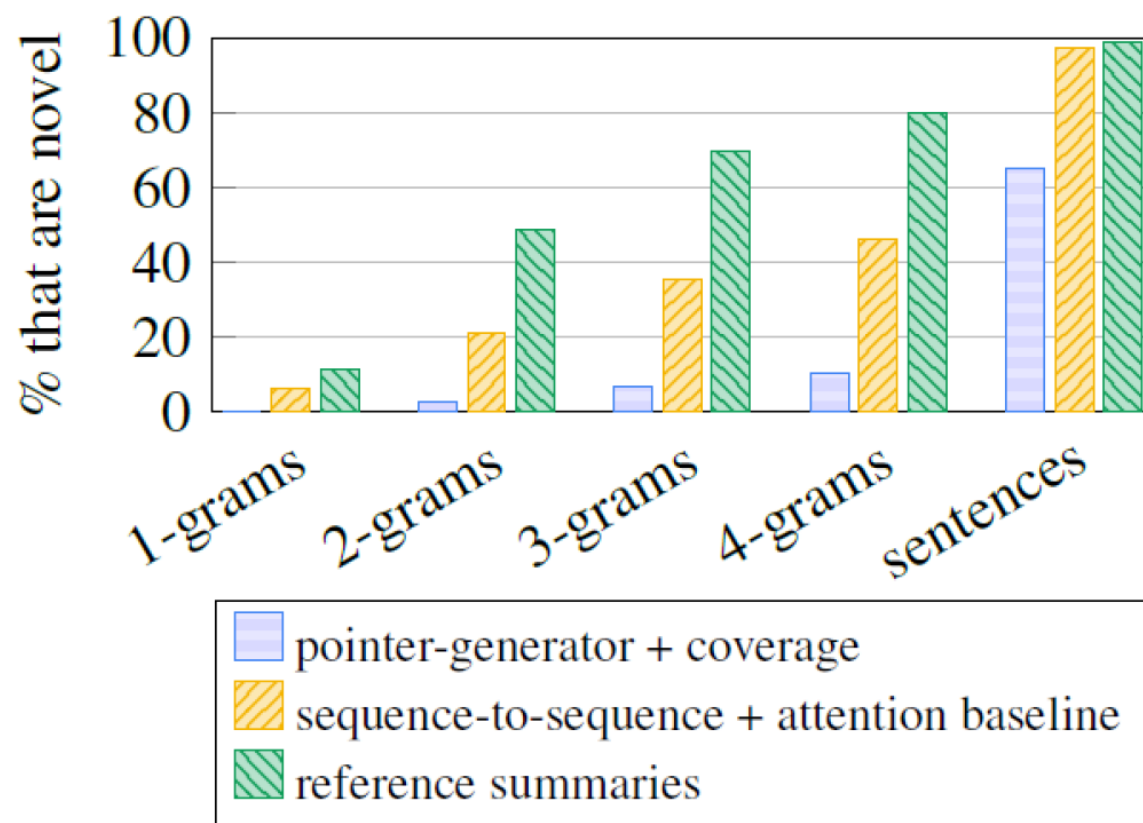
copies whole article sentences 35%