Neural Text Summarization

Piji Li

NLP Center, Tencent AI Lab
pijili@tencent.com

Paper Reading, Sep.6, 2018

Table of Contents

1. Introduction
2. Methods
3. Conclusion

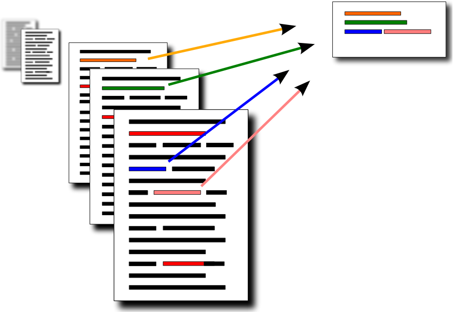
Table of Contents

1 Introduction

2 Methods

3 Conclusion

- The goal of automatic text summarization is to automatically produce a succinct summary, preserving the most important information for a single document or a set of documents about the same topic (event).



- Input:
  - Single-Document Summarization (SDS)
  - Multi-Document Summarization (MDS)

Introduction

Single-Document Summarization

Document

Cambodian leader Hun Sen on Friday rejected opposition parties ' demands for talks outside the country , accusing them of trying to `` internationalize " the political crisis .
Government and opposition parties have asked King Norodom Sihanouk to host a summit meeting after a series of post-election negotiations between the two opposition groups and Hun Sen 's party to form a new government failed .
Opposition leaders Prince Norodom Ranariddh and Sam Rainsy , citing Hun Sen 's threats to arrest opposition figures after two alleged attempts on his life , said they could not negotiate freely in Cambodia and called for talks at Sihanouk 's residence in Beijing .Hun Sen , however , rejected that .``
I would like to make it clear that all meetings related to Cambodian affairs must be conducted in the Kingdom of Cambodia , " Hun Sen told reporters after a Cabinet meeting on Friday .`` No-one should internationalize Cambodian affairs .
It is detrimental to the sovereignty of Cambodia , " he said .Hun Sen 's Cambodian People 's Party won 64 of the 122 parliamentary seats in July 's elections , short of the two-thirds majority needed to form a government on its own .Ranariddh and Sam Rainsy have charged that Hun Sen 's victory in the elections was achieved through widespread fraud .They have demanded a thorough investigation into their election complaints as a precondition for their cooperation in getting the national assembly moving and a new government formed ....

Summary

Cambodian government rejects opposition's call for talks abroad

Figure 1: Single-document summarization.

Introduction

Multi-Document Summarization

Documents

Fingerprints and photos of two men who boarded the doomed Malaysia Airlines passenger jet are being sent to U.S. authorities so they can be compared against records of known terrorists and criminals. The cause of the plane's disappearance has baffled investigators and they have not said that they believed that terrorism was involved, but they are also not ruling anything out. The investigation into the disappearance of the jetliner with 239 passengers and crew has centered so far around the fact that two passengers used passports stolen in Thailand from an Austrian and an Italian. The plane which left Kuala Lumpur, Malaysia, was headed for Beijing. Three of the passengers, one adult and two children, were American. ……

(CNN) -- A delegation of painters and calligraphers, a group of Buddhists returning from a religious gathering in Kuala Lumpur, a three-generation family, nine senior travelers and five toddlers. Most of the 227 passengers on board missing Malaysia Airlines Flight 370 were Chinese, according to the airline's flight manifest. The 12 missing crew members on the flight that disappeared early Saturday were Malaysian. The airline's list showed the passengers hailed from 14 countries, but later it was learned that two people named on the manifest -- an Austrian and an Italian -- whose passports had been stolen were not aboard the plane. The plane was carrying five children under 5 years old, the airline said. ……

⋮

Vietnamese aircraft spotted what they suspected was one of the doors belonging to the ill-fated Malaysia Airlines Flight MH370 on Sunday, as troubling questions emerged about how two passengers managed to board the Boeing 777 using stolen passports. The discovery comes as officials consider the possibility that the plane disintegrated mid-flight, a senior source told Reuters. The state-run Thanh Nien newspaper cited Lt. Gen. Vo Van Tuan, deputy chief of staff of Vietnam's army, as saying searchers in a low-flying plane had spotted an object suspected of being a door from the missing jet. It was found in waters about 56 miles south of Tho Chu island, in the same area where oil slicks were spotted Saturday. ……

Summary

Flight MH370, carrying 239 people vanished over the South China Sea in less than an hour after taking off from Kuala Lumpur, with two passengers boarded the Boeing 777 using stolen passports. Possible reasons could be an abrupt breakup of the plane or an act of terrorism. The government was determining the "true identities" of the passengers who used the stolen passports. Investigators were trying to determine the path of the plane by analysing civilian and military radar data while ships and aircraft from seven countries scouring the seas around Malaysia and south of Vietnam.

Figure 2: Multi-document summarization for the topic "Malaysia Airlines Disappearance".

- Input:
  - Single-Document Summarization (SDS)
  - Multi-Document Summarization (MDS)
- Output:
  - Extractive
  - Compressive
  - Abstractive
- Machine learning methods:
  - Supervised
  - Unsupervised

- Since 1950s:
  - Concept Weight (Luhn, 1958), Centroid (Radev et al., 2004), LexRank (Erkan and Radev, 2004), TextRank (Mihalcea and Tarau, 2004), Sparse Coding (He et al., 2012; Li et al., 2015)
  - Feature+Regression (Min et al., 2012; Wang et al., 2013)
- Most of the summarization methods are extractive.
- Abstractive summarization is full of challenges. Some indirect methods employ sentence fusing (Barzilay and McKeown, 2005) or phrase merging (Bing et al., 2015).
- The indirect strategies will do harm to the linguistic quality of the constructed sentences.

- Before the neural summarization era...silent
- 2012
- 2015 (Rush et al., 2015)

Table of Contents

1 Introduction

2 Methods

3 Conclusion

- Salience Detection (Words, Sentences)

- Liu, Yan, Sheng-hua Zhong, and Wenjie Li. "Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning." In AAAI. 2012.
- Denil, Misha, Alban Demiraj, Nal Kalchbrenner, Phil Blunsom, and Nando de Freitas. "Modelling, visualising and summarising documents with a single convolutional neural network." arXiv preprint arXiv:1406.3830 (2014).

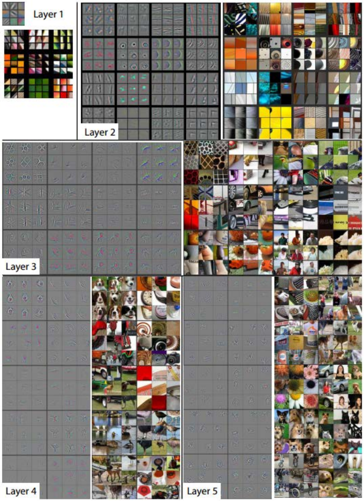


Figure 3: Visualization of Parameters.

- Since 1950s:
  - Concept Weight (Luhn, 1958), Centroid (Radev et al., 2004), LexRank (Erkan and Radev, 2004), TextRank (Mihalcea and Tarau, 2004), Sparse Coding (He et al., 2012; Li et al., 2015)
- **Bag-of-Words (BoWs)**

- Word2vec (Mikolov et al., 2013), Paragraph Vector (Le and Mikolov, 2014), RNN-Sent (Tang et al., 2015), CNN-Sent (Kim, 2014)
- Improve the performance of PageRank and Data Reconstruction based models.
- Works:
  - Kågebäck, Mikael, Olof Mogren, Nina Tahmasebi, and Devdatt Dubhashi. "**Extractive summarization using continuous vector space models**." In CVSC 2014.
  - Yin, Wenpeng, and Yulong Pei. "**Optimizing Sentence Modeling and Selection for Document Summarization**." In IJCAI 2015.
  - Li, Piji, Wai Lam, Lidong Bing, Weiwei Guo, and Hang Li. "**Cascaded attention based unsupervised information distillation for compressive summarization**." In EMNLP 2017.

- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "**Neural machine translation by jointly learning to align and translate**." arXiv preprint arXiv:1409.0473 (2014). (citation:4300+)

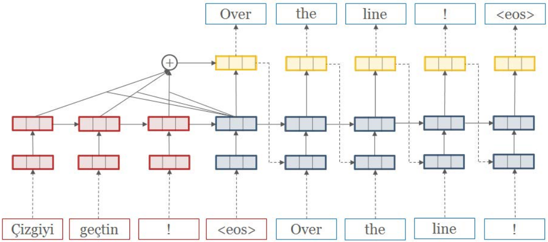


Figure 4: Attention-based seq2seq framework. Figure from OpenNMT (Klein et al., 2017)

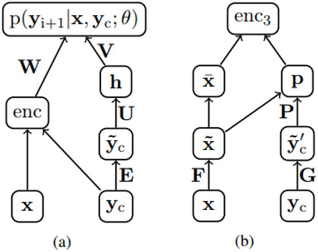
Methods

- 2015

Methods

A Neural Attention Model for Abstractive Sentence Summarization

- Rush, Alexander M., Sumit Chopra, and Jason Weston. "**A neural attention model for abstractive sentence summarization**." EMNLP (2015). (citation:570+)



$$
\begin{aligned}
\mathrm{enc}_3(\mathbf{x}, \mathbf{y}_c) &= \mathbf{p}^\top \bar{\mathbf{x}}, \\
\mathbf{p} &\propto \exp(\tilde{\mathbf{x}} \mathbf{P} \tilde{\mathbf{y}}_c'), \\
\tilde{\mathbf{x}} &= [\mathbf{F}\mathbf{x}_1, \ldots, \mathbf{F}\mathbf{x}_M], \\
\tilde{\mathbf{y}}_c' &= [\mathbf{G}\mathbf{y}_{i-C+1}, \ldots, \mathbf{G}\mathbf{y}_i], \\
\forall i \quad \bar{\mathbf{x}}_i &= \sum_{q=i-Q}^{i+Q} \tilde{\mathbf{x}}_i / Q.
\end{aligned}
$$

Figure 5: (a) NNLM decoder with additional encoder element. (b) Attention based encoder.

- Rush, Alexander M., Sumit Chopra, and Jason Weston. "**A neural attention model for abstractive sentence summarization**." EMNLP (2015). (citation:570+)

Model	DUC-2004 ROUGE-1	DUC-2004 ROUGE-2	DUC-2004 ROUGE-L	Gigaword ROUGE-1	Gigaword ROUGE-2	Gigaword ROUGE-L	Gigaword Ext. %
IR	11.06	1.67	9.67	16.91	5.55	15.58	29.2
PREFIX	22.43	6.49	19.65	23.14	8.25	21.73	100
COMPRESS	19.77	4.02	17.30	19.63	5.13	18.28	100
W&L	22	6	17	-	-	-	-
TOPIARY	25.12	6.46	20.12	-	-	-	-
MOSES+	26.50	8.13	22.85	28.77	12.10	26.44	70.5
ABS	26.55	7.06	22.05	30.88	12.22	27.77	85.4
ABS+	28.18	8.49	23.81	31.00	12.65	28.34	91.5
REFERENCE	29.21	8.38	24.46	-	-	-	45.6

Table 1: Experimental results on the main summary tasks on various ROUGE metrics . Baseline models are described in detail in Section 7.2. We report the percentage of tokens in the summary that also appear in the input for Gigaword as Ext %.

- Hu, Baotian, Qingcai Chen, and Fangze Zhu. "**LCSTS: A Large Scale Chinese Short Text Summarization Dataset**." EMNLP (2015). (citation:49)

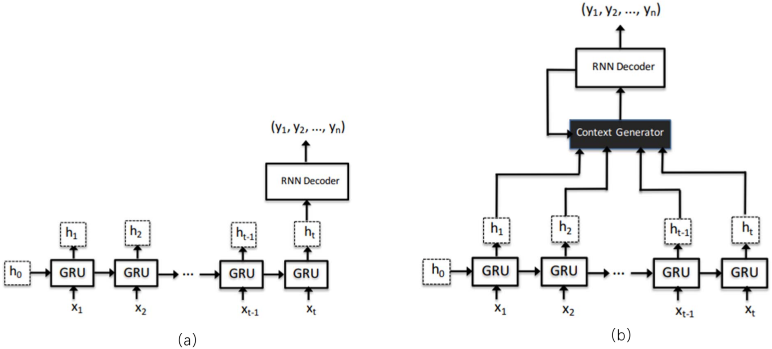


Figure 6: (a) Encoder-Decoder. (b) Attention based Decoder.

Methods

LCSTS: A Large Scale Chinese Short Text Summarization Dataset

- Hu, Baotian, Qingcai Chen, and Fangze Zhu. "**LCSTS: A Large Scale Chinese Short Text Summarization Dataset**." EMNLP (2015). (citation:49)

【江西高考被曝替考 有关考生已被警方控制】人民日报记者吴齐强消息，江西高考被曝光替考，7日中午江西省教育厅发布消息称，接到有人组织替考的举报后，江西省教育厅、江西省教育考试院立即部署南昌市教育考试院，联合南昌市警方开展调查核实，有关考生已被警方控制。调查进展情况将及时向社会公布。

Figure 1: A Weibo Posted by People's Daily.

model	data	R-1	R-2	R-L
RNN	Word	0.177	0.085	0.158
	Char	0.215	0.089	0.186
RNN context	Word	0.268	0.161	0.241
	Char	**0.299**	**0.174**	**0.272**

Table 2: The experiment result: "Word" and "Char" denote the word-based and character-based input respectively.

Short Text: 9月3日，总部位于日内瓦的世界经济论坛发布了《2014-2015年全球竞争力报告》，瑞士连续六年位居榜首，成为全球最具竞争力的国家，新加坡和美国分列第二位和第三位。中国排名第28位，在金砖国家中排名最高。

On September 3, the Geneva-based World Economic Forum released "The Global Competitiveness Report 2014-2015". Switzerland topped the list for six consecutive years , becoming the world 's most competitive country. Singapore and the United States are in the second and third place respectively. China is in the 28th place, ranking highest among the BRIC countries.

Human: 全球竞争力排行榜中国居28位居金砖国家首位

The Global competitiveness ranking list, China is in the 28th place, the highest among BRIC countries.

RNN+Char: 全球竞争力报告：瑞士居第六位居第26位第三位居第26位第六位居第26位第六位居第26位第六位居第26位

"The Global Competitiveness Report" : Switzerland ranks 6th in 26th ranks 3rd ranks 26th 6th ranks 26th place 6th place ranks 26th place 6th place ranks 26th place

RNN+Word: 2014年全球竞争力排名：瑞士第一北京第第第第名单第第第名单第68位世界第第第名单第68位

2014 Global competitiveness ranking: Switzerland ranks first place, Beijing ranks ranks ranks ranks list ranks ranks ranks list ranks 68th world ranks ranks ranks list ranks 68th .

RNN+Context+Char: 全球竞争力排名：中国排名第28位金砖国最高

The Global competitiveness ranking list: China is in the 28th place which is highest among the BRIC countries.

RNN+Context+Word: 2014-2015年全球竞争力报告：瑞士连续6年居榜首中国居28位(首/3———访榜首)中国排名第28位

"The Global Competitiveness Report 2014-2015" : Switzerland topped the list for six consecutive years, China is in the 28th place (*can not be translated*) China ranks 28th.

Figure 7: An example of the generated summaries.

- Lopyrev, Konstantin. "**Generating news headlines with recurrent neural networks**." arXiv preprint arXiv:1512.01712 (2015). (citation:28)
- Investigations of several NMT models.

Methods

- 2016

- Chopra, Sumit, Michael Auli, and Alexander M. Rush. "**Abstractive sentence summarization with attentive recurrent neural networks**." NAACL, pp. 93-98. 2016. (citation:138)

- Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Ça glar Gulçehre, and Bing Xiang. "**Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond**." CoNLL 2016 (2016): 280. (citation:183)
- 3 pages version in Feb. 2016.
- Many tricks (nmt, copy, coverage, hierarchical, external knowledge).

Methods

Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond

- Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Ça glar Gulçehre, and Bing Xiang. "**Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond**." CoNLL 2016 (2016): 280. (citation:183)

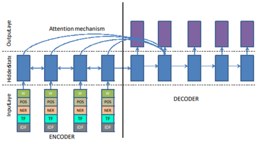


Figure 1: Feature-rich-encoder: We use one embedding vector each for POS, NER tags and discretized TF and IDF values, which are concatenated together with word-based embeddings as input to the encoder.

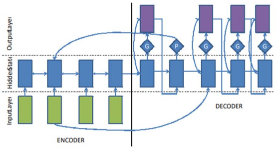


Figure 2: Switching generator/pointer model:

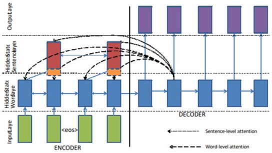


Figure 3: Hierarchical encoder with hierarchical attention

- Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Ça glar Gulçehre, and Bing Xiang. "**Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond**." CoNLL 2016 (2016): 280. (citation:183)

#	Model name	Rouge-1	Rouge-2	Rouge-L	Src. copy rate (%)
	Full length F1 on our internal test set				
1	words-lvt2k-1sent	34.97	17.17	32.70	75.85
2	words-lvt2k-2sent	35.73	17.38	33.25	79.54
3	words-lvt2k-2sent-hieratt	36.05	18.17	33.52	78.52
4	feats-lvt2k-2sent	35.90	17.57	33.38	78.92
5	feats-lvt2k-2sent-ptr	*__36.40__	**17.77**	*__33.71__	78.70
	Full length F1 on the test set used by (Rush et al., 2015)				
6	ABS+ (Rush et al., 2015)	29.78	11.89	26.97	91.50
7	words-lvt2k-1sent	32.67	15.59	30.64	74.57
8	RAS-Elman (Chopra et al., 2016)	33.78	15.97	31.15	
9	words-lvt5k-1sent	*__35.30__	**16.64**	*__32.62__	

Table 1: Performance comparison of various models. '*' indicates statistical significance of the corresponding model with respect to the baseline model on its dataset as given by the 95% confidence interval in the official Rouge script. We report statistical significance only for the best performing models. 'src. copy rate' for the reference data on our validation sample is 45%. Please refer to Section 4 for explanation of notation.

- Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Ça glar Gulçehre, and Bing Xiang. "**Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond**." CoNLL 2016 (2016): 280. (citation:183)

Model	Rouge-1	Rouge-2	Rouge-L
words-lvt2k	32.49	11.84	29.47
words-lvt2k-hieratt	32.75	12.21	29.01
words-lvt2k-temp-att	***35.46**	***13.30**	***32.65**

Table 3: Performance of various models on CNN/Daily Mail test set using full-length Rouge-F1 metric. Bold faced numbers indicate best performing system.

- Nallapati, Ramesh, Bowen Zhou, Cicero dos Santos, Ça glar Gulçehre, and Bing Xiang. "**Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond**." CoNLL 2016 (2016): 280. (citation:183)

Model	Rouge-1	Rouge-2	Rouge-L
words-lvt2k	32.49	11.84	29.47
words-lvt2k-hieratt	32.75	12.21	29.01
words-lvt2k-temp-att	*****35.46**	*****13.30**	*****32.65**

Table 3: Performance of various models on CNN/Daily Mail test set using full-length Rouge-F1 metric. Bold faced numbers indicate best performing system.

- OOV
- Extraction

Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly. "**Pointer networks**." In NIPS, pp. 2692-2700. 2015. (citation:352)

- Gulcehre, Caglar, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio. "**Pointing the Unknown Words**." In ACL, vol. 1, pp. 140-149. 2016. (citation:126)
- Gu, Jiatao, Zhengdong Lu, Hang Li, and Victor OK Li. "**Incorporating Copying Mechanism in Sequence-to-Sequence Learning**." In ACL, vol. 1, pp. 1631-1640. 2016. (citation:192)

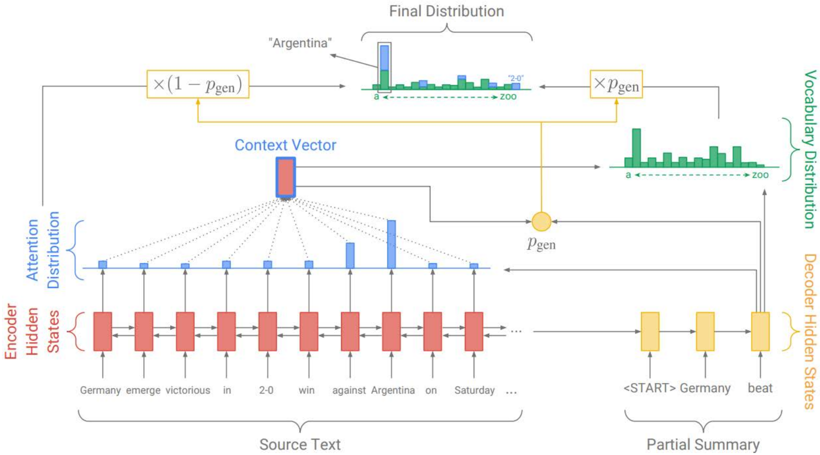


Figure 7: Pointer-generator model. (See et al., 2017)

	Rouge-1	Rouge-2	Rouge-L
NMT + lvt	34.87	16.54	32.27
NMT + lvt + PS	**35.19**	**16.66**	**32.51**

Table 1: Results on Gigaword Corpus when pointers are used for UNKs in the training data, using Rouge-F1 as the evaluation metric.

Models		ROUGE scores on LCSTS (%) R-1	R-2	R-L
RNN (Hu et al., 2015)	+C	21.5	8.9	18.6
	+W	17.7	8.5	15.8
RNN context (Hu et al., 2015)	+C	29.9	17.4	27.2
	+W	26.8	16.1	24.1
COPYNET	+C	**34.4**	**21.6**	**31.3**
	+W	**35.0**	**22.3**	**32.0**

Table 3: Testing performance of LCSTS, where "RNN" is canonical Enc-Dec, and "RNN context" its attentive variant.

Methods

Why Coverage?

- Diversity

Tu, Zhaopeng, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. "**Modeling Coverage for Neural Machine Translation**." In ACL 2016. (citation:187)

- Chen, Qian, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. "**Distraction-based neural networks for modeling documents**." In IJCAI 2016. (citation:28)

- Chen, Qian, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. "**Distraction-based neural networks for modeling documents**." In IJCAI 2016. (citation:28)

$$c_t = \tanh(W_c c'_t - U_c \sum_{j=1}^{t-1} c_j)$$

$$\alpha'_{t,i} = v_a^T \tanh(W_a s'_t + U_a h_i - b_a \sum_{j=1}^{t-1} \alpha_{j,i})$$

Figure 8: Operation of coverage mechanism.

- Chen, Qian, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. "**Distraction-based neural networks for modeling documents**." In IJCAI 2016. (citation:28)

System	Rouge-1	Rouge-2	Rouge-L
Luhn	23.2	7.2	15.5
Edmundson	24.5	8.2	16.7
LSA	21.2	6.2	14.0
Lex-rank	26.1	**9.6**	17.7
Text-rank	23.3	7.7	15.8
Sum-basic	22.9	5.5	14.8
KL-sum	20.7	5.9	13.7
Uni-GRU	18.4	4.8	14.3
Bi-GRU	19.5	5.2	15.0
+Two-level out	20.2	5.9	15.7
+UNK replace	21.3	6.3	16.4
+Distraction M1	22.2	6.5	16.7
+Distraction M2	24.4	7.7	17.8
+Distraction M3	**27.1**	8.2	**18.7**

Table 2: Results on the CNN dataset.

System	Rouge-1	Rouge-2	Rouge-L
[Hu *et al.*, 2015]	29.9	17.4	27.2
Uni-GRU	32.1	19.9	29.4
Bi-GRU	33.2	20.8	30.5
+Two-level Att.	35.2	22.6	32.5
+UNK replace	35.2	22.6	32.5
+Distraction	35.2	22.6	32.5

Table 4: Results on the LCSTS dataset.

Methods

More Works in 2016[1]

- Cheng, Jianpeng, and Mirella Lapata. "Neural Summarization by Extracting Sentences and Words." In ACL, 2016. (citation:108)
- Cao, Ziqiang, Wenjie Li, Sujian Li, Furu Wei, and Yanran Li. "AttSum: Joint Learning of Focusing and Summarization with Neural Attention." In COLING, 2016.
- Zeng, Wenyuan, Wenjie Luo, Sanja Fidler, and Raquel Urtasun. "Efficient summarization with read-again and copy mechanism." arXiv preprint arXiv:1611.03382 (2016).
- Miao, Yishu, and Phil Blunsom. "Language as a Latent Variable: Discrete Generative Models for Sentence Compression." In EMNLP. 2016.
- ...

[1] https://github.com/lipiji/App-DL#text-summarization

Methods

- 2017

- Inspirations from the traditional summarization methods.

Methods

- Nallapati, Ramesh, Feifei Zhai, and Bowen Zhou. "**SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents**." In AAAI, pp. 3075-3081. 2017. (citation:58)

- Tan, Jiwei, Xiaojun Wan, and Jianguo Xiao. "**Abstractive document summarization with a graph-based attentional neural model**." In ACL 2017. (citation:24)
- ACL Outstanding Paper.

- Tan, Jiwei, Xiaojun Wan, and Jianguo Xiao. "**Abstractive document summarization with a graph-based attentional neural model**." In ACL 2017. (citation:24)

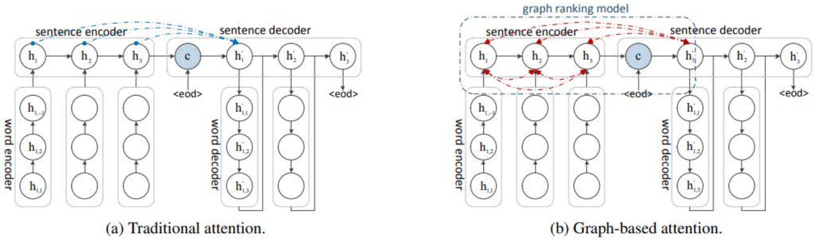


Figure 1: Hierarchical encoder-decoder framework and comparison of the attention mechanisms.

- Zhou, Qingyu, Nan Yang, Furu Wei, and Ming Zhou. "**Selective Encoding for Abstractive Sentence Summarization**." In ACL 2017. (citation:24)

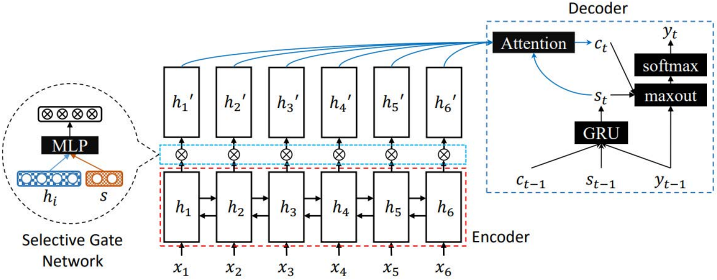


Figure 2: Overview of the **S**elective **E**ncoding for **A**bstractive **S**entence **S**ummarization (**SEASS**).

Methods

- Recall the Copy and Coverage Mechanism in 2016.

- See, Abigail, Peter J. Liu, and Christopher D. Manning. "**Get To The Point: Summarization with Pointer-Generator Networks**." In ACL 2017. (citation:114)
- Writing? Figures?

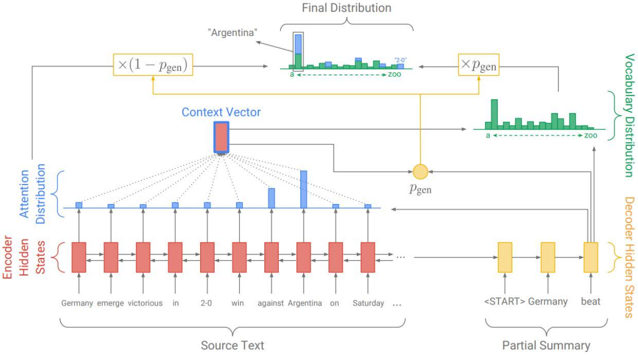


Figure 9: Pointer-Generator Networks.

Methods

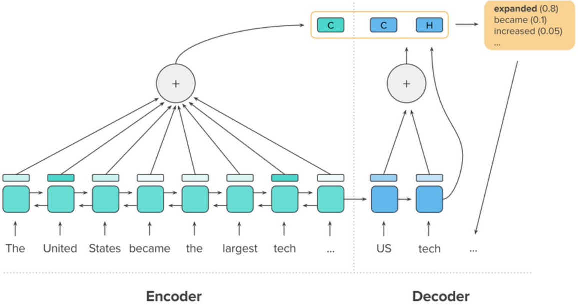
- Reinforcement Learning.

- Paulus, Romain, Caiming Xiong, and Richard Socher. "**A deep reinforced model for abstractive summarization**." arXiv preprint arXiv:1705.04304 (2017). (citation:107)

Methods

A deep reinforced model for abstractive summarization

- Intra-attention modeling.
- Reinforced learning trick.



$$L_{ml} = -\sum_{t=1}^{n'} \log p(y_t^*|y_1^*, \ldots, y_{t-1}^*, x)$$

$$L_{rl} = (r(\hat{y}) - r(y^s)) \sum_{t=1}^{n'} \log p(y_t^s|y_1^s, \ldots, y_{t-1}^s, x)$$

$$L_{mixed} = \gamma L_{rl} + (1-\gamma) L_{ml}.$$

- Paulus, Romain, Caiming Xiong, and Richard Socher. "**A deep reinforced model for abstractive summarization**." arXiv preprint arXiv:1705.04304 (2017). (citation:107)

Model	ROUGE-1	ROUGE-2	ROUGE-L
Lead-3 (Nallapati et al., 2017)	39.2	15.7	35.5
SummaRuNNer (Nallapati et al., 2017)	39.6	16.2	35.3
words-lvt2k-temp-att (Nallapati et al., 2016)	35.46	13.30	32.65
ML, no intra-attention	37.86	14.69	34.99
ML, with intra-attention	38.30	14.81	35.49
RL, with intra-attention	**41.16**	15.75	**39.08**
ML+RL, with intra-attention	39.87	**15.82**	36.90

Table 1: Quantitative results for various models on the CNN/Daily Mail test dataset

Methods

- 2018

- Celikyilmaz, Asli, Antoine Bosselut, Xiaodong He, and Yejin Choi. "**Deep Communicating Agents for Abstractive Summarization**." In NAACL 2018.
- Wu, Yuxiang, and Baotian Hu. "**Learning to Extract Coherent Summary via Deep Reinforcement Learning**." In AAAI 2018.
- Wang, Li, Junlin Yao, Yunzhe Tao, Li Zhong, Wei Liu, and Qiang Du. "**A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization**." In IJCAI 2018.
- Chen, Yen-Chun, and Mohit Bansal. "**Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting**." arXiv preprint arXiv:1805.11080 (2018).
- Keneshloo, Yaser, Tian Shi, Chandan K. Reddy, and Naren Ramakrishnan. "**Deep Reinforcement Learning For Sequence to Sequence Models**." arXiv preprint arXiv:1805.09461 (2018).

- Wang, Li, Junlin Yao, Yunzhe Tao, Li Zhong, Wei Liu, and Qiang Du. "**A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization**." In IJCAI 2018.

- Wojciech Kryściński, Romain Paulus, Caiming Xiong, Richard Socher. "Improving Abstraction in Text Summarization ." arXiv preprint arXiv:1808.07913 (2018).
- Zhang, Xingxing, Mirella Lapata, Furu Wei, and Ming Zhou. "Neural Latent Extractive Document Summarization." arXiv preprint arXiv:1808.07187 (2018).
- Sebastian Gehrmann, Yuntian Deng, Alexander M. Rush. "Bottom-Up Abstractive Summarization." arXiv preprint arXiv:1808.10792 (2018).

Methods

More:

- https://github.com/lipiji/App-DL#text-summarization

Table of Contents

1 Introduction

2 Methods

3 Conclusion

Conclusion

- Challenges:
  - Long text abstractive summarization.
  - Abstractive multi-document summarization.

Thanks a lot!

Q & A

References I

Regina Barzilay and Kathleen R McKeown. Sentence fusion for multi-document news summarization. *Computational Linguistics*, 31(3):297–328, 2005.

Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca Passonneau. Abstractive multi-document summarization via phrase selection and merging. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 1587–1597, 2015.

Günes Erkan and Dragomir R Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22:457–479, 2004.

References II

Zhanying He, Chun Chen, Jiajun Bu, Can Wang, Lijun Zhang, Deng Cai, and Xiaofei He. Document summarization based on data reconstruction. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 620–626. AAAI Press, 2012.

Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751, 2014.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M Rush. Opennmt: Open-source toolkit for neural machine translation. *arXiv preprint arXiv:1701.02810*, 2017.

Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188–1196, 2014.

References III

Piji Li, Lidong Bing, Wai Lam, Hang Li, and Yi Liao. Reader-aware multi-document summarization via sparse coding. In *The 24th International Joint Conference on Artificial Intelligence*, pages 1270–1276, 2015.

Hans Peter Luhn. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165, 1958.

Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Ziheng Lin Min, Yen Kan Chew, and Lim Tan. Exploiting category-specific information for multi-document summarization. *The 21th International Conference on Computational Linguistics (COLING)*, pages 2903–2108, 2012.

References IV

Dragomir R Radev, Hongyan Jing, Małgorzata Styś, and Daniel Tam. Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6):919–938, 2004.

Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, 2015.

Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1073–1083, 2017.

Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 1422–1432, 2015.

References V

Lu Wang, Hema Raghavan, Vittorio Castelli, Radu Florian, and Claire Cardie. A sentence compression based framework to query-focused multi-document summarization. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1384–1394, 2013.