ACL19 Summarization

Xiachong Feng

Papers

- Multi-Document Summarization
- Scientific Paper Summarization
- Pre-train Based Summarization
- Other Papers

Paper	Conference
Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking	ACL19
Self-Supervised Learning for Contextualized Extractive Summarization	ACL19
BiSET: Bi-directional Selective Encoding with Template for Abstractive Summarization	ACL19
Multi-News: a Large-Scale Multi-Document Summarization Dataset and Abstractive Hierarchical Model	ACL19
Hierarchical Transformers for Multi-Document Summarization	ACL19
HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization	ACL19
HIGHRES: Highlight-based Reference-less Evaluation of Summarization	ACL19
TALKSUMM: A Dataset and Scalable Annotation Method for Scientific Paper Summarization Based on Conference Talks	ACL19
BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization	ACL19
Searching for Effective Neural Extractive Summarization: What Works and What's Next	ACL19
Generating Summaries with Topic Templates and Structured Convolutional Decoders	ACL19
Self-Supervised Learning for Contextualized Extractive Summarization	ACL19

Overview

- Total 30 (3 student workshop)
  - Extractive : 4
  - Abstractive : 9
  - Unsupervised ： 3

Dataset

- Multi-News: a Large-Scale **Multi-Document** Summarization Dataset and Abstractive Hierarchical Model
- BIG**PATENT**: A Large-Scale Dataset for Abstractive and Coherent Summarization
- TalkSumm: A Dataset and Scalable Annotation Method for **Scientific Paper** Summarization Based on Conference Talks

Cross-lingual

- Zero-Shot **Cross-Lingual** Abstractive Sentence Summarization through Teaching Generation and Attention
  - Mingming Yin, Xiangyu Duan, Min Zhang, Boxing Chen and Weihua Luo

Multi-Document

- Multi-News: a Large-Scale **Multi-Document** Summarization Dataset and Abstractive Hierarchical Model
- Hierarchical Transformers for **Multi-Document** Summarization
  - Yang Liu and Mirella Lapata
- Improving the Similarity Measure of Determinantal Point Processes for Extractive **MultiDocument** Summarization
  - Sangwoo Cho, Logan Lebanoff, Hassan Foroosh and Fei Liu

Multi-Modal

- **Multimodal** Abstractive Summarization for How2 Videos
  - Shruti Palaskar, Jindřich Libovický, Spandana Gella and Florian Metze
- Keep Meeting Summaries on Topic: Abstractive **Multi-Modal** Meeting Summarization
  - Manling Li, Lingyu Zhang, Heng Ji and Richard J. Radke

Unsupervised

- Simple **Unsupervised** Summarization by Contextual Matching
  - Jiawei Zhou and Alexander Rush
- **Unsupervised** Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking
  - Masaru Isonuma, Junichiro Mori and Ichiro Sakata
- Sentence Centrality Revisited for **Unsupervised** Summarization
  - Hao Zheng and Mirella Lapata

Multi-Document

Multi-Document Summarization

- GENERATING WIKIPEDIA BY SUMMARIZING LONG SEQUENCES ***ICLR18***
- Hierarchical Transformers for Multi-Document Summarization ***ACL19***
- Multi-News: a Large-Scale Multi-Document Summarization Dataset and Abstractive Hierarchical Model ***ACL19***

- Graph-based Neural Multi-Document Summarization ***CoNLL17***

Multi-Doc Summarization Dataset

- DUC
- WikiSum *(ICLR18)*
- Multi-News *(ACL19)*

DUC

- Document Understanding Conferences (DUC)
- DUC 2001, 2002, 2003 and 2004 containing 30, 59, 30 and 50 clusters of nearly 10 documents each respectively.
- Trained on DUC 2001 and 2002, validated on 2003, and tested on 2004

	DUC'01	DUC'02	DUC'03	DUC'04
# of Clusters	30	59	30	50
# of Documents	309	567	298	500
# of Sentences	24498	16090	7721	13270
Vocabulary Size	28188	22174	13248	18036
Summary Length	100 words	100 words	100 words	665 Bytes

WikiSum

- GENERATING WIKIPEDIA BY SUMMARIZING LONG SEQUENCES ***ICLR18***
- **Input:**
  - Title of a Wikipedia article
  - Collection of source documents
    - Webpages cited in the References section of the Wikipedia article
    - The top 10 search results returned by Google
- **Output:**
  - Wikipedia article's first section
- Train/Dev/Test
  - 1865750, 233252, and 232998

Multi-News

- Multi-News: a Large-Scale Multi-Document Summarization Dataset and Abstractive Hierarchical Model ***ACL19***
- Large-scale MDS news dataset
- https://www.newser.com/
- 56,216 articles-summary pairs.
- Each summary is professionally written by editors and includes links to the original articles cited.

# of source	Frequency	# of source	Frequency
2	23,894	7	382
3	12,707	8	209
4	5,022	9	89
5	1,873	10	33
6	763		

Dataset	# pairs	# words (doc)	# sents (docs)	# words (summary)	# sents (summary)	vocab size
Multi-News	44,972/5,622/5,622	2,103.49	82.73	263.66	9.97	666,515

Multi-News

Source 1

Meng Wanzhou, Huawei's chief financial officer and deputy chair, was arrested in Vancouver on 1 December. Details of the arrest have not been released...

Source 2

A Chinese foreign ministry spokesman said on Thursday that Beijing had separately called on the US and Canada to "clarify the reasons for the detention "immediately and "immediately release the detained person ". The spokesman...

Source 3

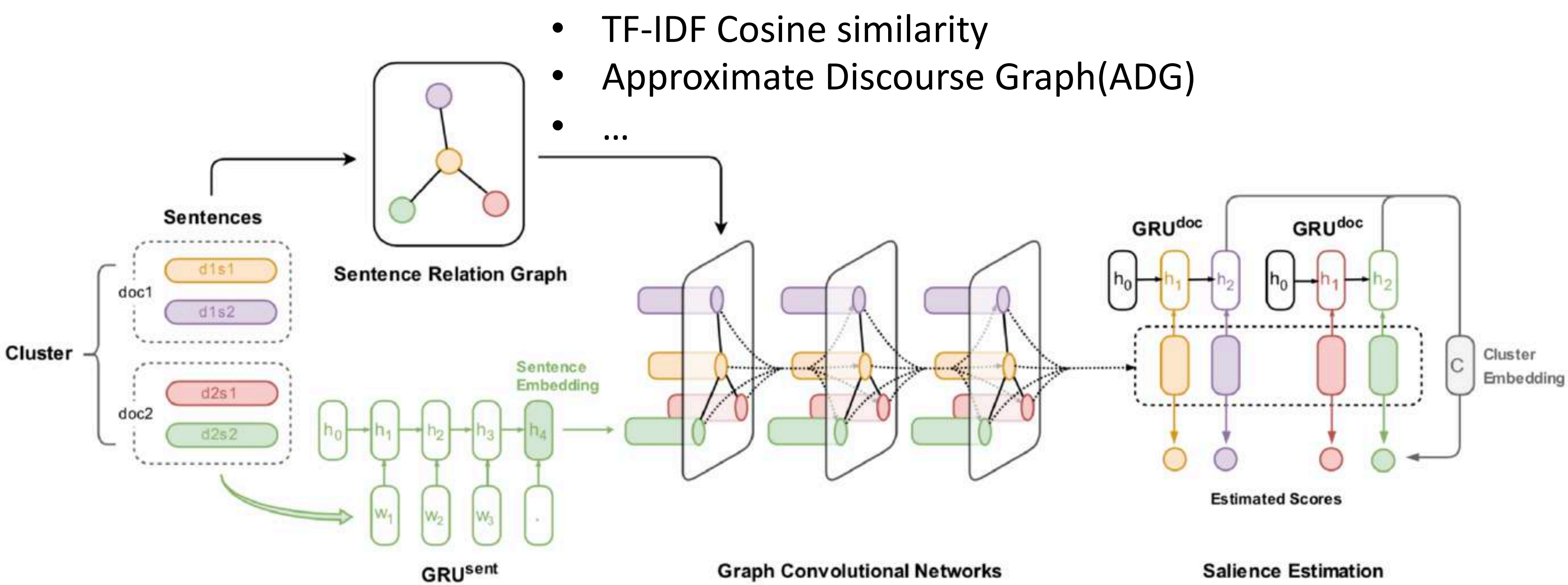
Canadian officials have arrested Meng Wanzhou, the chief financial officer and deputy chair of the board for the Chinese tech giant Huawei,...Meng was arrested in Vancouver on Saturday and is being sought for extradition by the United States. A bail hearing has been set for Friday...

Summary

...Canadian authorities say she was being sought for extradition to the US, where the company is being investigated for possible violation of sanctions against Iran. Canada's justice department said Meng was arrested in Vancouver on Dec. 1... China's embassy in Ottawa released a statement.. "The Chinese side has lodged stern representations with the US and Canadian side, and urged them to immediately correct the wrongdoing "and restore Meng's freedom, the statement said...

Relations Among Documents

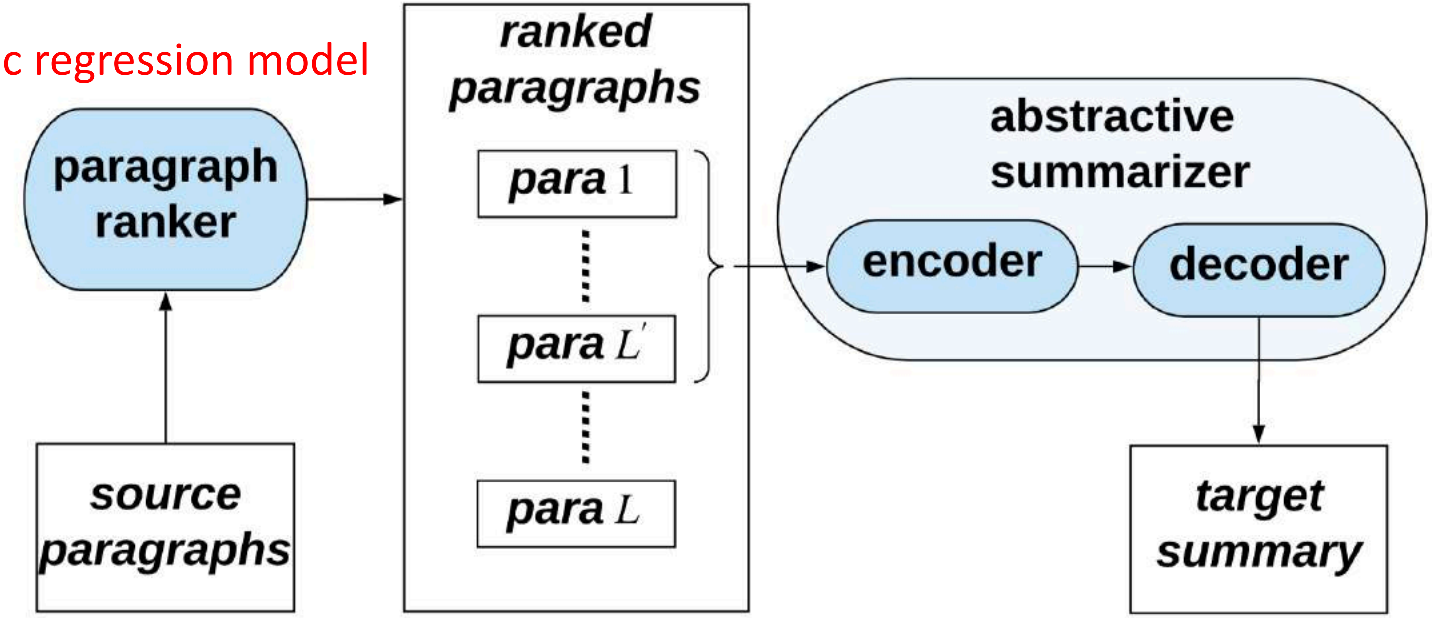
- The importance of considering relations among sentences in multi-document summarization.



Hierarchical Transformers for Multi-Document Summarization

- ACL19
- WikiSum Dataset

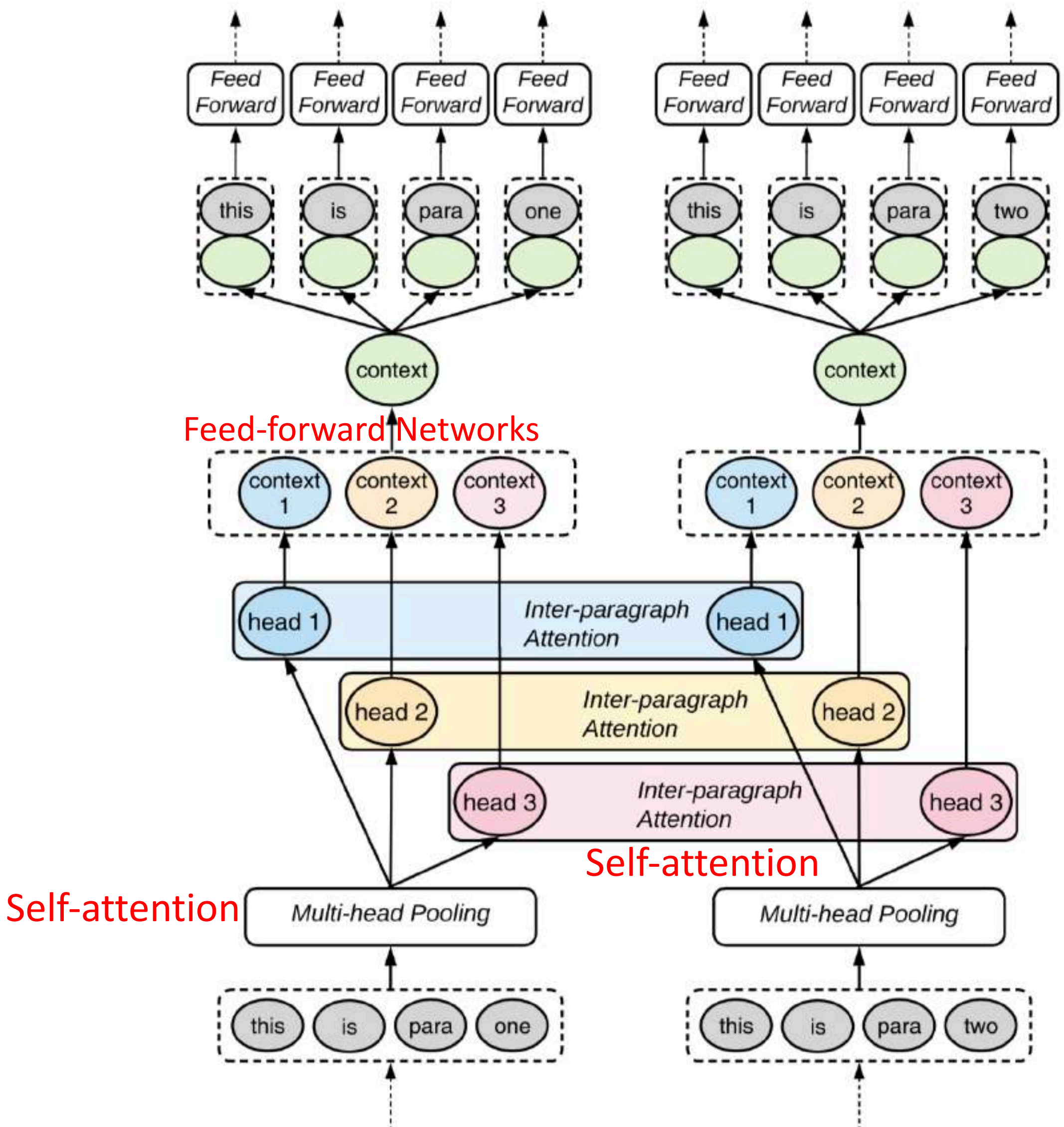
Logistic regression model



Hierarchical Transformers

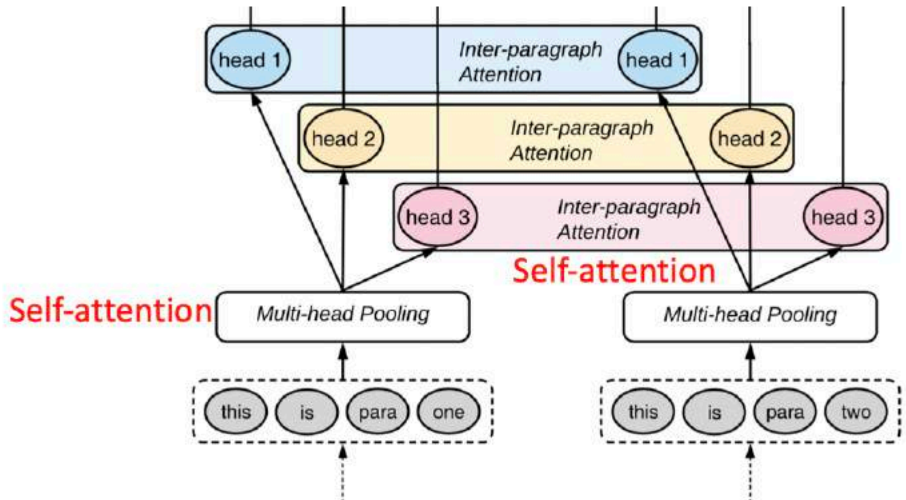
- **Input**
  - Word embedding
  - Paragraph position embedding
  - Sentence position embedding
- **Local Transformer Layer**
  - Encode contextual information for tokens within each paragraph
- **Global Transformer Layer**
  - Exchange information across multiple paragraphs

Hierarchical Transformers-Encoder



Graph-informed Attention

- Cosine similarities based on tf-idf
- Discourse relations



HT (1,600 tokens)	**40.82**	**25.99**	35.08
HT (1,600 tokens) + Similarity Graph	40.80	25.95	35.08
HT (1,600 tokens) + Discourse Graph	40.81	25.95	**35.24**

Scientific Paper

Scientific Paper Summarization

- **TALKSUMM**: A Dataset and Scalable Annotation Method for Scientific Paper Summarization Based on Conference Talks ***ACL19***
- **ScisummNet**: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks ***AAAI19***

Dataset

- TALKSUMM ***(ACL19)***
- Scisumm ***(AAAI19)***

TALKSUMM

- Automatically generate **extractive** content-based summaries for scientific papers based on video talks

Title: Split and Rephrase: Better Evaluation and Stronger Baselines (Aharoni and Goldberg, 2018)

Paper: Processing long, complex sentences is challenging. This is true either for humans in various circumstances or in NLP tasks like parsing and machine translation . An automatic system capable of breaking a complex sentence into several simple sentences that convey the same meaning is very appealing . A recent work by Narayan et al. (2017) introduced a dataset, evaluation method and baseline systems for the task, naming it Split-and Rephrase . The dataset includes 1,066,115 instances mapping a single complex sentence to a sequence of sentences that express the same meaning, together with RDF triples that describe their semantics. They considered two . . . Indeed, feeding the model with examples containing entities alone without any facts about them causes it to output perfectly phrased but unsupported facts (Table 3). Digging further, we find that 99% of the simple sentences (more than 89% of the unique ones) in the validation and test sets also appear in the training set, which coupled with the good memorization capabilities of SEQ2SEQ models and the relatively small number of distinct simple sentences helps to explain the high BLEU score . To aid further research on the task, we propose a more challenging split of the data . We also establish a stronger baseline by extending the SEQ2SEQ approach with a copy mechanism, which was shown . . . We encourage future work on the split-and-rephrase task to use our new data split or the v1.0 split instead of the original one.

Talk transcript: let's begin with the motivation so processing long complex sentences is a hard task this is true for arguments like children people with reading disabilities second language learners but this is also true for sentence level and NLP systems , for example previous work show that dependency parsers degrade performance when they're introduced with longer and longer sentences, in a similar result was shown for neural machine translation , where neural machine translation systems introduced with longer sentences starting degrading performance, the question rising here is can we automatically break a complex sentence into several simple ones while preserving the meaning or the semantics and this can be a useful component in NLP pipelines . For example, the split and rephrase task was introduced in the last EMNLP by Narayan, Gardent and Shimarina, where they introduced a dataset, an evaluation method and baseline models for this task. The task definition can be taking a complex sentence and breaking it into several simple ones with the same meaning . For example, . . . semantics units in the source sentence and then rephrasing those units into a single sentences on the target site. In this work we first show the simple neural models seem to perform very well on the original benchmark, but this is only due to memorization of the training set , we propose a more challenging data split for the task to discourage this memorization and we perform automatic evaluation in error analysis on the new benchmark showing that the task is still very far from being solved.

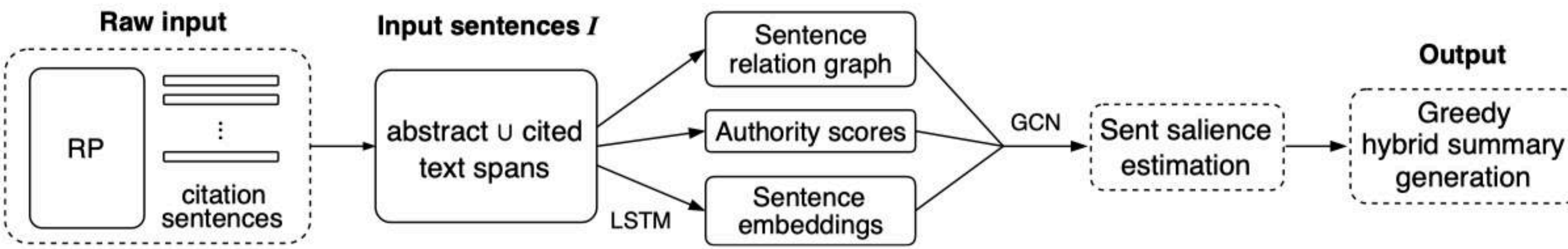
TALKSUMM

- NLP and ML
  - ACL, NAACL, EMNLP, SIGDIAL (2015-2018), and ICML (2017-2018).
- Create a new dataset, that contains 1716 summaries for papers from several computer science conferences
- HMM
  - The sequence of spoken words is the output sequence.
  - Each hidden state of the HMM corresponds to a single paper sentence.
- Four training sets, two with fixed-length summaries (150 and 250 words), and two with fixed ratio between summary and paper lengths (0.3 and 0.4).

Scisumm

- ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks ***AAAI19***
- 1,000 most cited papers in the ACL Anthology Network (AAN)
- Summary : not only the major points **highlighted by the authors (abstract)** but also the views offered by the **scientific community**
- **Input:**
  - Reference paper
  - Citation sentence
- **Output:**
  - Summary
    - Read its abstract and incoming citation sentences to create a gold summary. Without reading the whole text

Scisumm



Raw input
RP
citation sentences
Input sentences I
abstract ∪ cited text spans
LSTM
Sentence relation graph
Authority scores
Sentence embeddings
GCN
Sent salience estimation
Output
Greedy hybrid summary generation

Pre-train Based

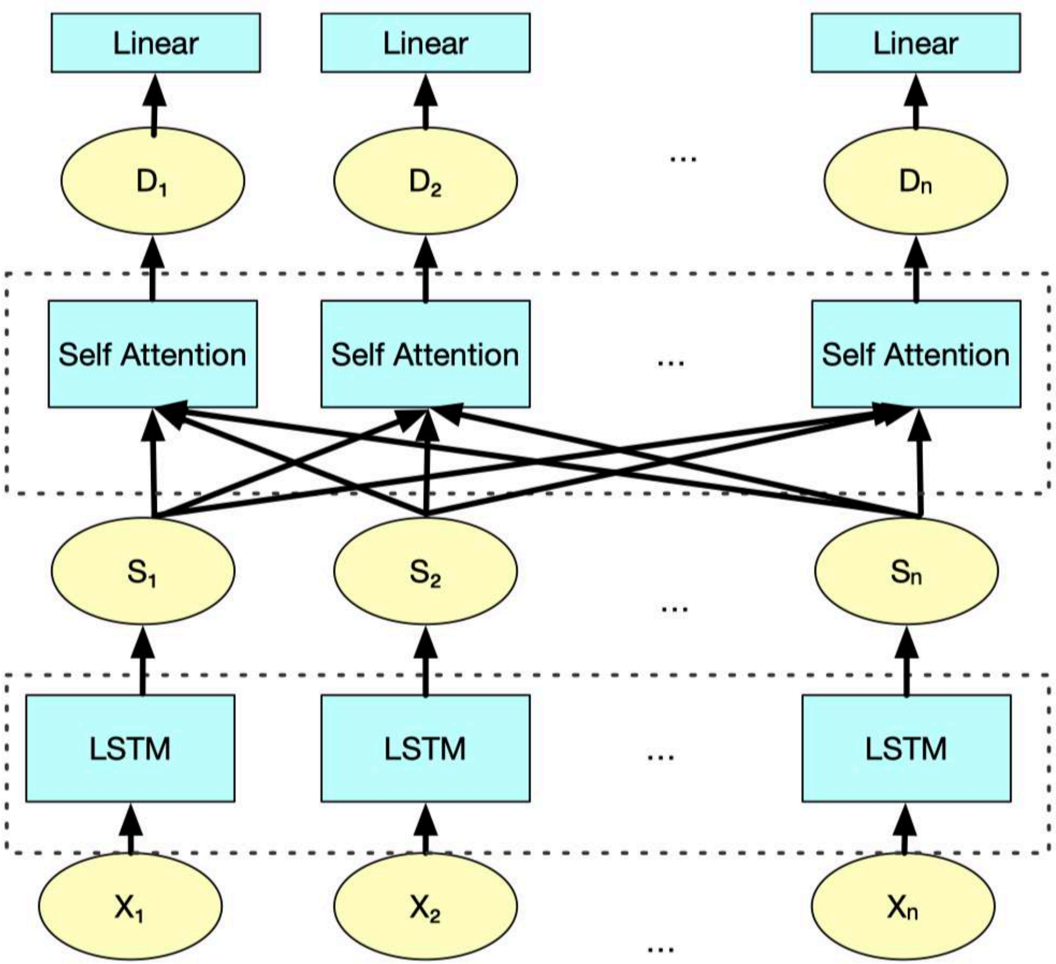
Pre-train Based Summarization

- Self-Supervised Learning for Contextualized Extractive Summarization ***ACL19***
- HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization ***ACL19***

Self-Supervised Learning

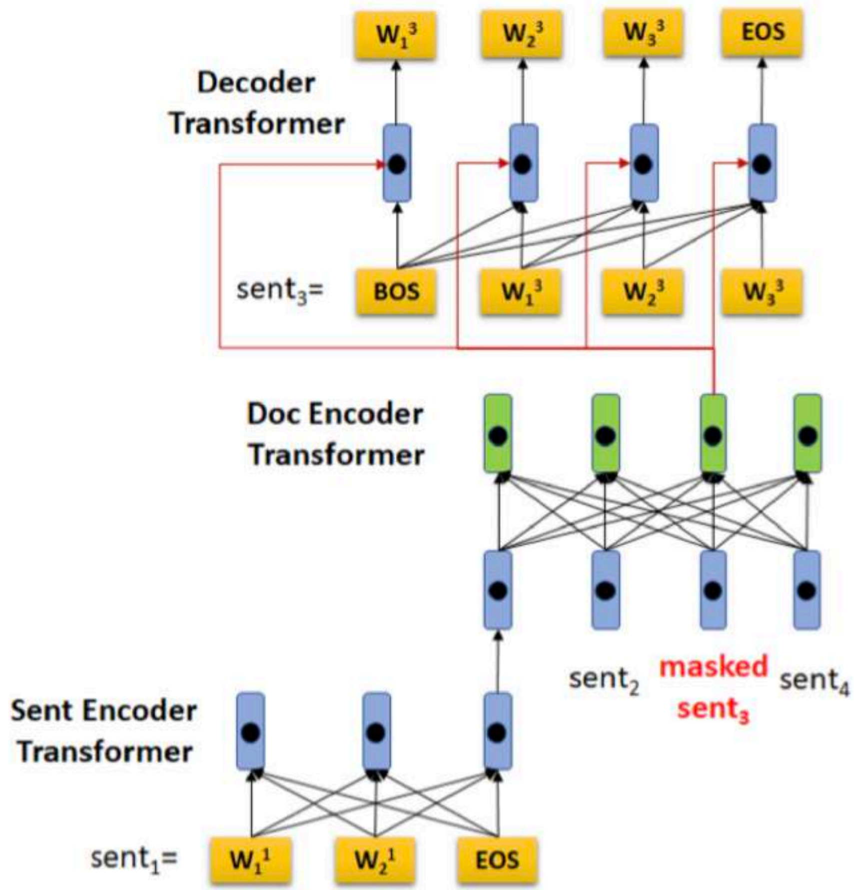
- Self-Supervised Learning for Contextualized Extractive Summarization ***ACL19***
- The **Mask task** randomly masks some sentences and predicts the missing sentence from a candidate pool
- The **Replace task** randomly replaces some sentences with sentences from other documents and predicts if a sentence is replaced.
- The **Switch task** switches some sentences within the same document and predicts if a sentence is switched.

Self-Supervised Learning

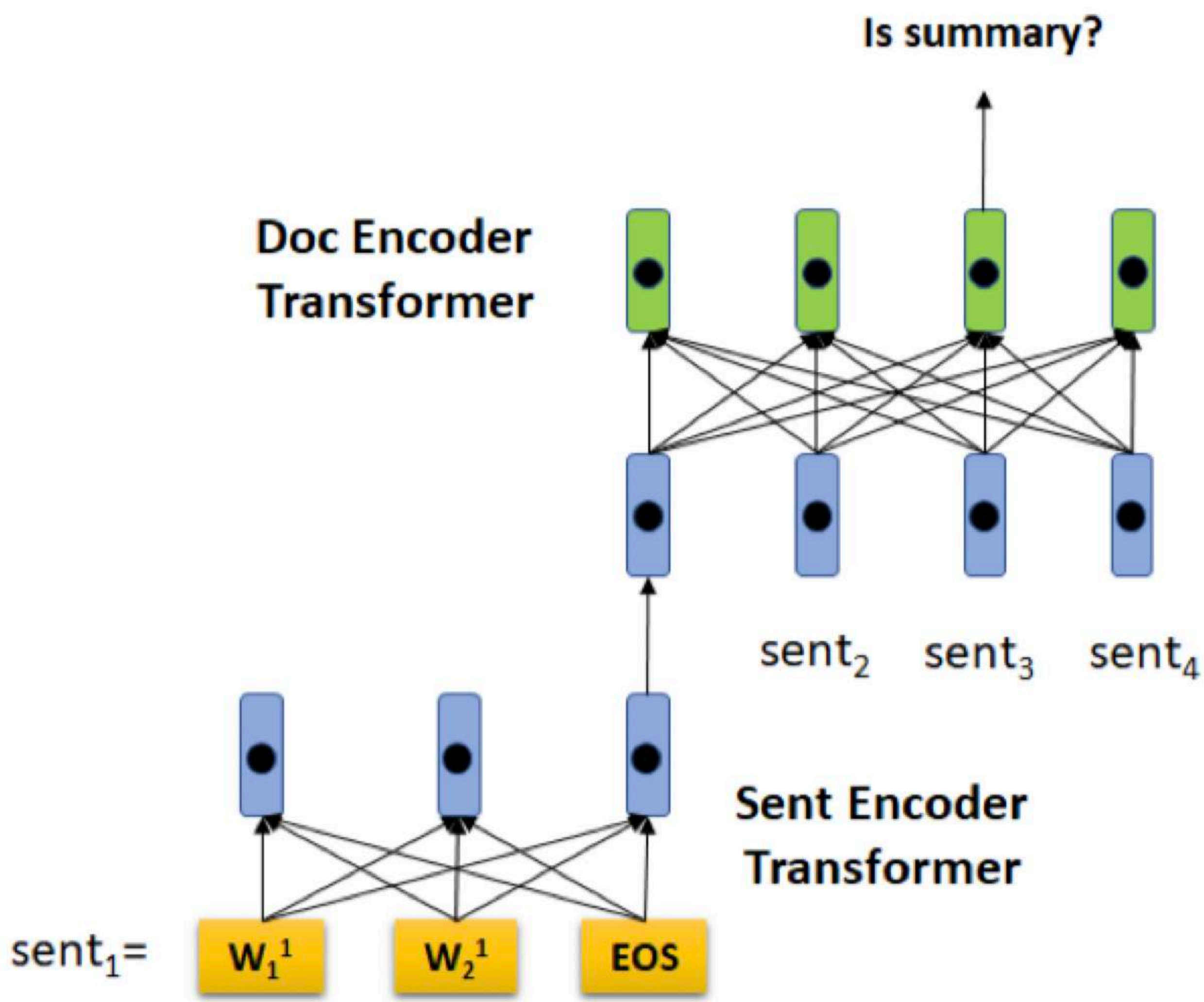


HIBERT

- HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization ***ACL19***



HIBERT



Is summary?
Doc Encoder
Transformer
sent$_2$
sent$_3$
sent$_4$
Sent Encoder
Transformer
sent$_1$=
W$_1^1$
W$_2^1$
EOS

Others

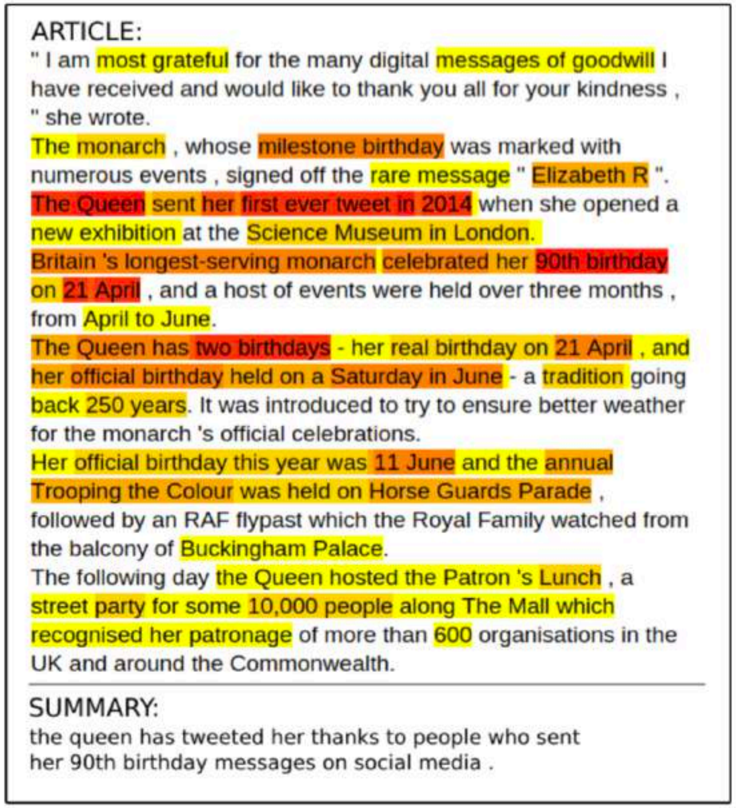
1. BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization ***ACL19***
2. HIGHRES: Highlight-based Reference-less Evaluation of Summarization ***ACL19***
3. Searching for Effective Neural Extractive Summarization: What Works and What‘s Next ***ACL19***
4. BiSET: Bi-directional Selective Encoding with Template for Abstractive Summarization ***ACL19***
5. Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking ***ACL19***

BIGPATENT

- BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization ***ACL19***
- 1.3 million records of U.S. patent documents（专利文献） along with human written abstractive summaries
- **Patent documents**
    - Title, authors, abstract, claims of the invention and the description text.
- **Core**
    - Summaries contain a richer discourse structure with more recurring entities
    - Salient content is evenly distributed in the input
    - Lesser and shorter extractive fragments are present in the summaries.

HIGHRES

- HIGHRES: Highlight-based Reference-less Evaluation of Summarization ***ACL19***
- Human Evaluation Framework


ARTICLE:
" I am most grateful for the many digital messages of goodwill I have received and would like to thank you all for your kindness , " she wrote.
The monarch , whose milestone birthday was marked with numerous events , signed off the rare message " Elizabeth R ".
The Queen sent her first ever tweet in 2014 when she opened a new exhibition at the Science Museum in London.
Britain 's longest-serving monarch celebrated her 90th birthday on 21 April , and a host of events were held over three months , from April to June.
The Queen has two birthdays - her real birthday on 21 April , and her official birthday held on a Saturday in June - a tradition going back 250 years. It was introduced to try to ensure better weather for the monarch 's official celebrations.
Her official birthday this year was 11 June and the annual Trooping the Colour was held on Horse Guards Parade , followed by an RAF flypast which the Royal Family watched from the balcony of Buckingham Palace.
The following day the Queen hosted the Patron 's Lunch , a street party for some 10,000 people along The Mall which recognised her patronage of more than 600 organisations in the UK and around the Commonwealth.

SUMMARY:
the queen has tweeted her thanks to people who sent her 90th birthday messages on social media .

HIGHRES

- **Highlight Annotation**
  - From single words to complete sentences or even paragraphs.
  - Limit in the number of words to K

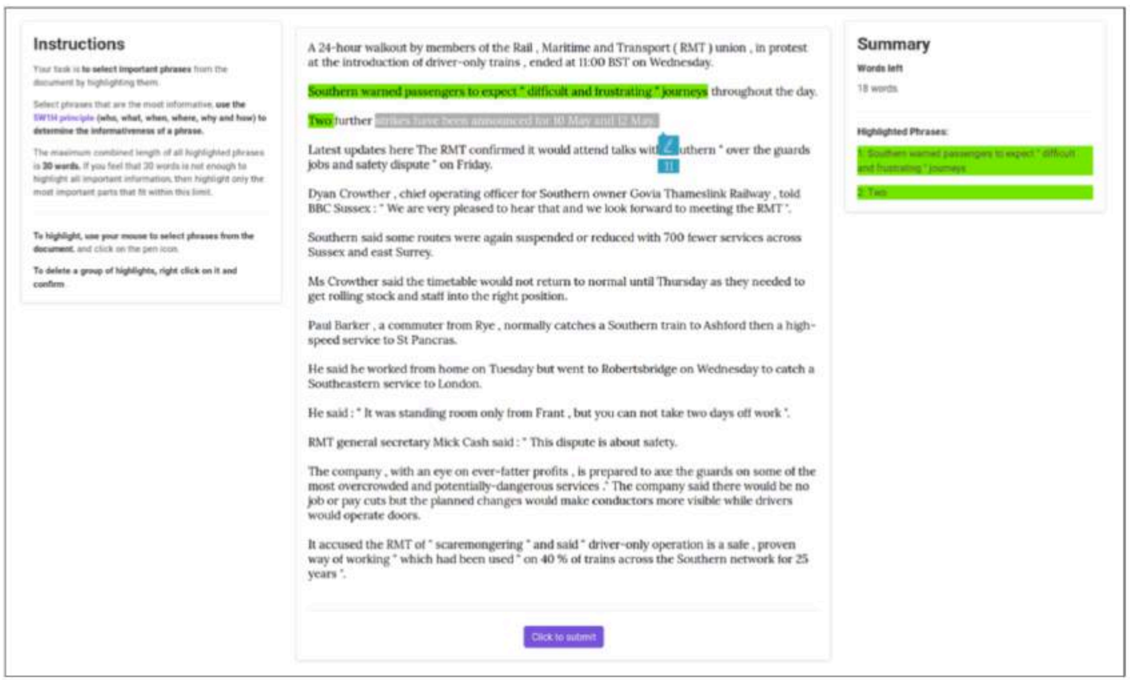


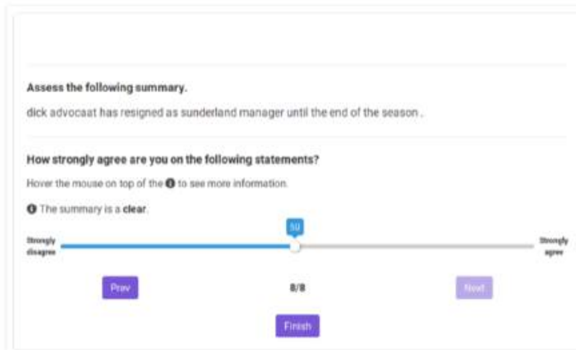
Figure 2: The UI for highlight annotation. Judges are given an article and asked to highlight words or phrases that are important in the article.

HIGHRES

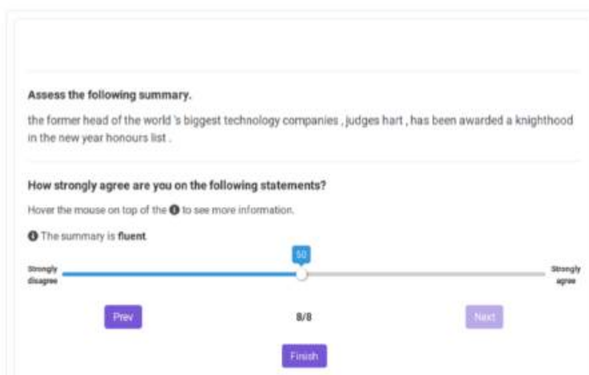
- **Highlight-based Content Evaluation**
  - **Given**:document that has been highlighted using heatmap coloring and a summary to assess.
  - **Recall (content coverage):** All important information is present in the summary (1-100)
  - **Precision (informativeness):** Only important information is in the summary. (1-100)

HIGHRES

- **Clarity**
  - Each judge is asked whether the summary is easy to be understood



- **Fluency**
  - Each judge is asked whether the summary sounds natural and has no grammatical problems.



HIGHRES

- **Highlight-based ROUGE Evaluation**
  - N-grams are weighted by the number of times they were highlighted.

HIGHRES Framework

1. **Recall (content coverage)**
2. **Precision (informativeness)**
3. Clarity
4. Fluency
5. **Highlight-based ROUGE Evaluation**

Experimental

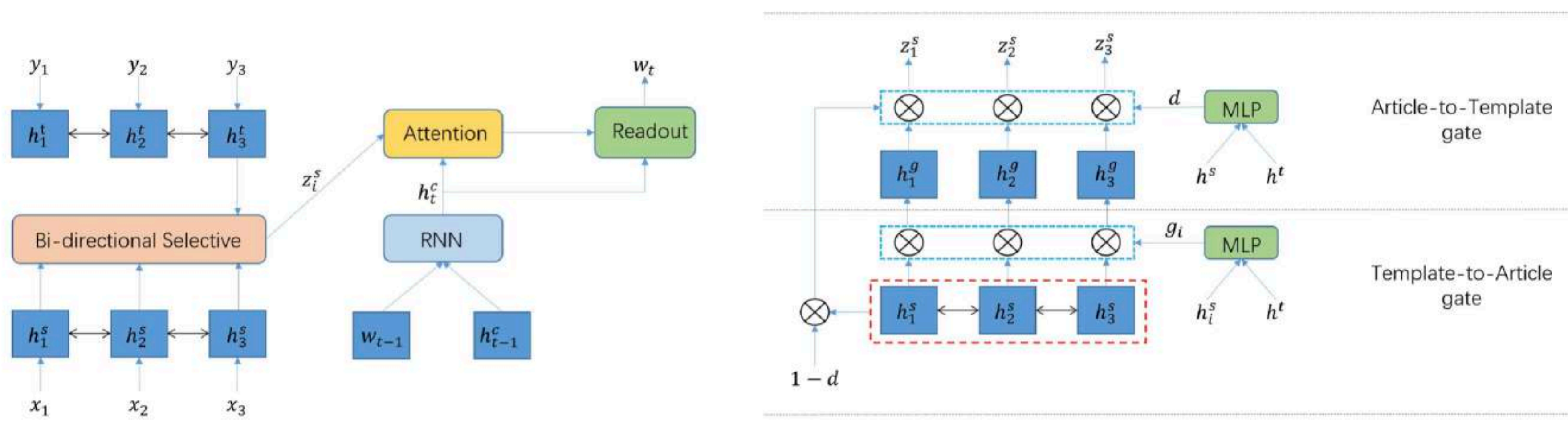
- Searching for Effective Neural Extractive Summarization: What Works and What's Next ***ACL19***

Conclusion

1. Auto-regressive is better than Non auto-regressive.
2. Pre-trained model and Reinforcement learning can further boost performance.
3. Transformer is more robust.

BiSET

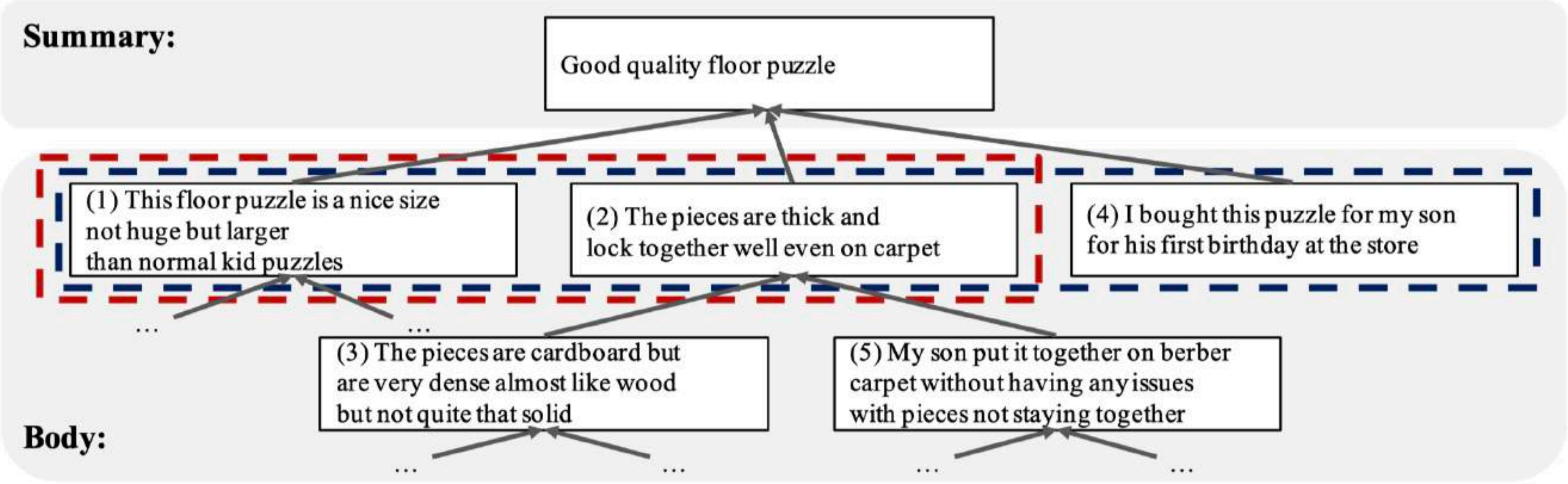
- BiSET: Bi-directional Selective Encoding with Template for Abstractive Summarization ***ACL19***
- ***Re3sum*** *(ACL18)* ***+ Co-attention***



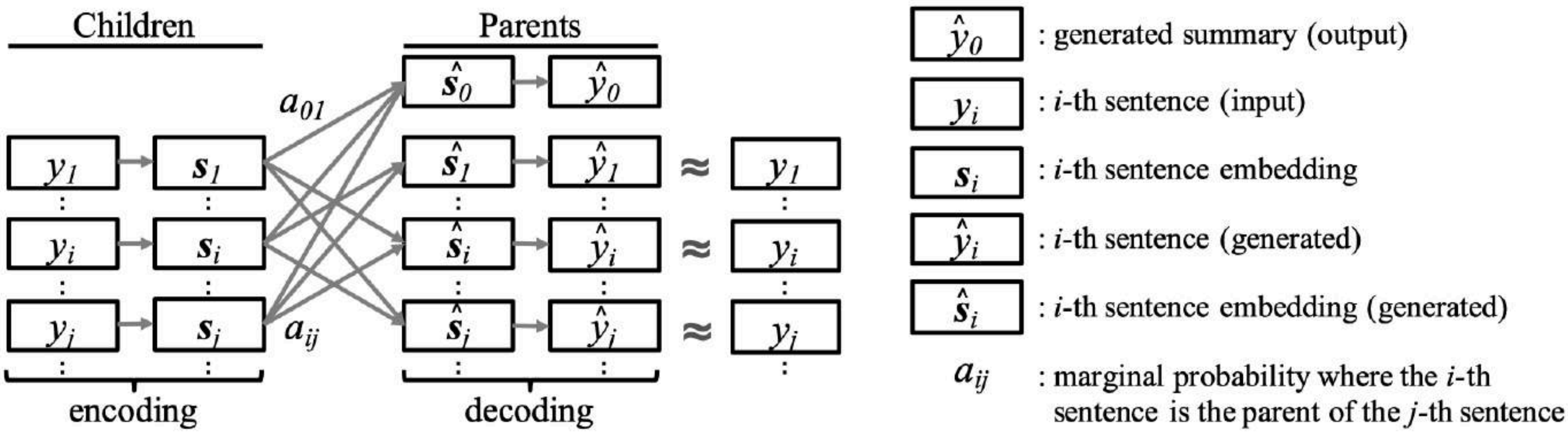
(a) (b)

Unsupervised

- Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking ***ACL19***



Unsupervised



$$\sum_{i=0}^{n} a_{ik} = 1$$

Multi-News: a Large-Scale Multi-Document Summarization Dataset and Abstractive Hierarchical Model **MDS Summarization dataset; News domain; 56,216;**
TALKSUMM: A Dataset and Scalable Annotation Method for Scientific Paper Summarization Based on Conference Talks **Extractive; Scientific paper; Video; NLP&ML domain;**
BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization **Patent doamin; Abstractive; Less lead bias**
Hierarchical Transformers for Multi-Document Summarization **Explicit and implicit graph modeling**
HIGHRES: Highlight-based Reference-less Evaluation of Summarization **Human Evaluation Framework**
Searching for Effective Neural Extractive Summarization: What Works and What's Next **Auto-regressive; Transformer; Pre-trained model; Reinforcememt learning**
BiSET: Bi-directional Selective Encoding with Template for Abstractive Summarization **Template; Retrive; Rerank; Co-attention**
Self-Supervised Learning for Contextualized Extractive Summarization **Mask; Replace; Switch**
HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization **Mask sentence; Decode the sentence**
Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking **Unsupervised; Discourse**

Thanks!