From Standard Summarization to New Tasks and Beyond: Summarization with Manifold Information

Shen Gao and Rui Yan

shengao@pku.edu.cn

Outline

1. Introduction to summarization with manifold information
    1. Traditional summarization
    2. New summarization task
    3. Challenges and problems
2. Background knowledge
    1. Concepts, problem formulation, and task statements
    2. Deep learning for summarization
3. Summarization by incorporating document structure
    1. Long document summarization
    2. Timeline summarization
    3. Dialog summarization
    4. Academic paper summarization
    5. Movie Summarization
4. Summarization by incorporating additional knowledge
    1. Reader-aware summarization
    2. Template-based summarization
    3. Multi-modal summarization
    4. Opinion Summarization
5. Recent trends
    1. Multi-modal summarization
    2. Long document summarization
    3. Dialog summarization
6. Summary

Target Audience

- Our target audiences are researchers and practitioners with some deep learning and text process background
- Our target audiences are interested in new summarization task and the technologies behind the prosperity of real-world summarization application in industry and academia.
- They would like to learn how to build a summarization system with state-of-the-art technologies.

Introduction

- Two types of text summarization
  - Summarizes a plain text
  - Generating summary with manifold information
- *New summarization* tasks aim to produce a better and appropriate summary by incorporating manifold information in many real-world applications.

Task of Traditional summarization

- Very simple and general
- Input: a plain text document
- Output: a short dense text describe the main idea of the input document

New summarization task

- Different with traditional summarization task
- Using structured document as input
- Leveraging other knowledge source as additional input
- These new summarization task can better adapt to real-world summarization applications

Challenges and problems

- How to understand the semantic meanings of the text with structure?
- How to incorporate additional knowledge when summarizing documents?

Background: Deep Learning for Summarization

- Extractive Summarization
  - **Sequence Labeling** uses an RNN to read the sentences only once
  - **Encoder-Decoder** uses two RNN to encode the passage and decode the sentence pointer.
  - **Reinforcement learning** method directly optimize the ROUGE score
  - **Pretraining** techniques employ the language model pre-training model
  - **Graph Model** contains additional nodes which act as the intermediary between sentences and enrich the cross-sentence relations
- Abstractive Summarization
  - **Sequence-to-sequence** based text generation methods
  - **Copy mechanism** directly copy the OOV words
  - **Selective encoding** encode the important semantic parts and ignore the trivial parts.
  - **Pretraining** techniques employ the language model pre-training model
  - **Contrastive Learning** bridge the gap between the *learning objective* and *evaluation metrics*

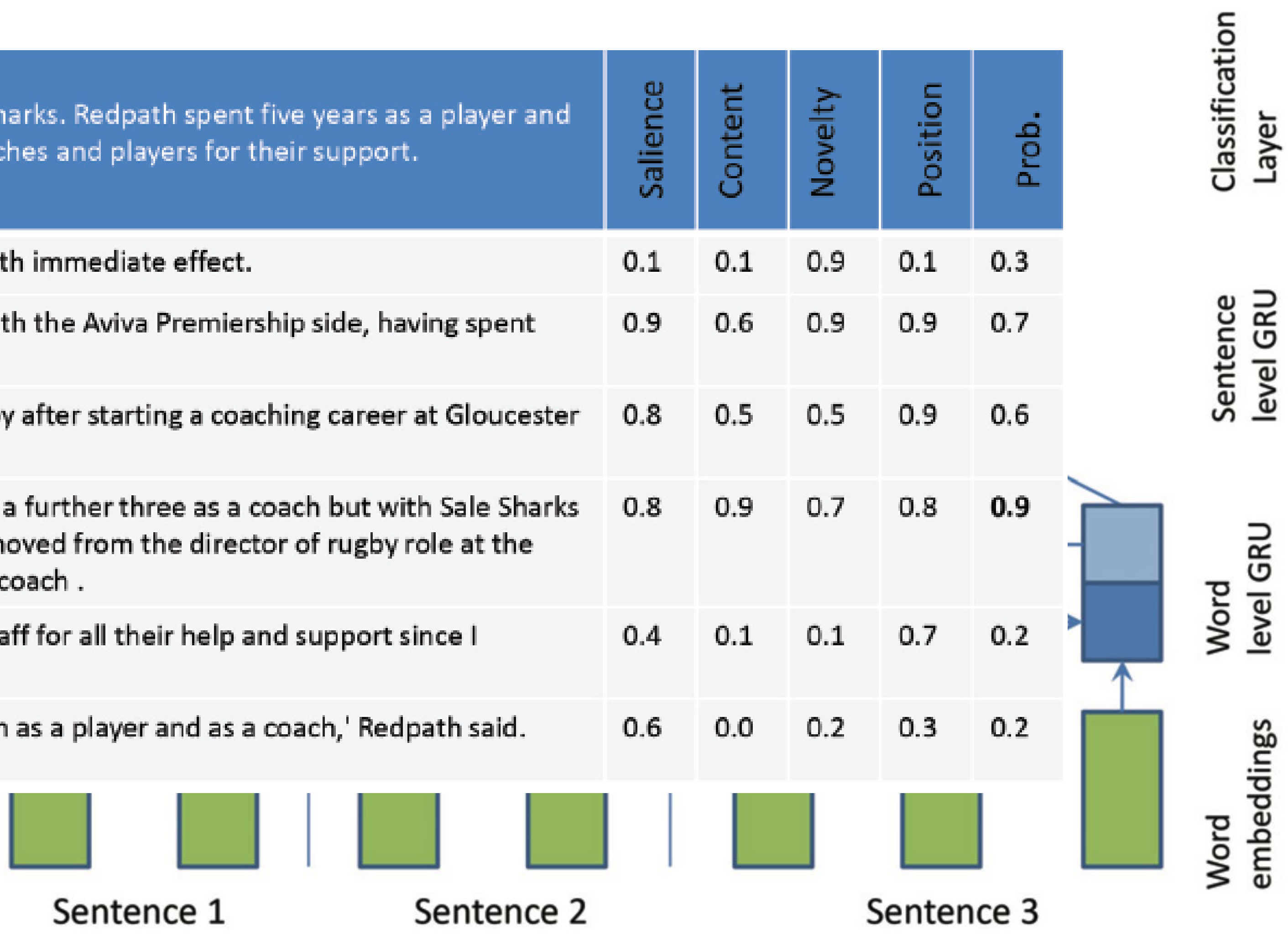
Datasets

- CNNDM
- WikiSum
- BIGPATENT
- Newsroom
- WikiHow
- XSUM

Extractive Summarization

- Tr
  su
  qu
  pr
- Se
  fe
  in
  sa
  cc
- A
  made

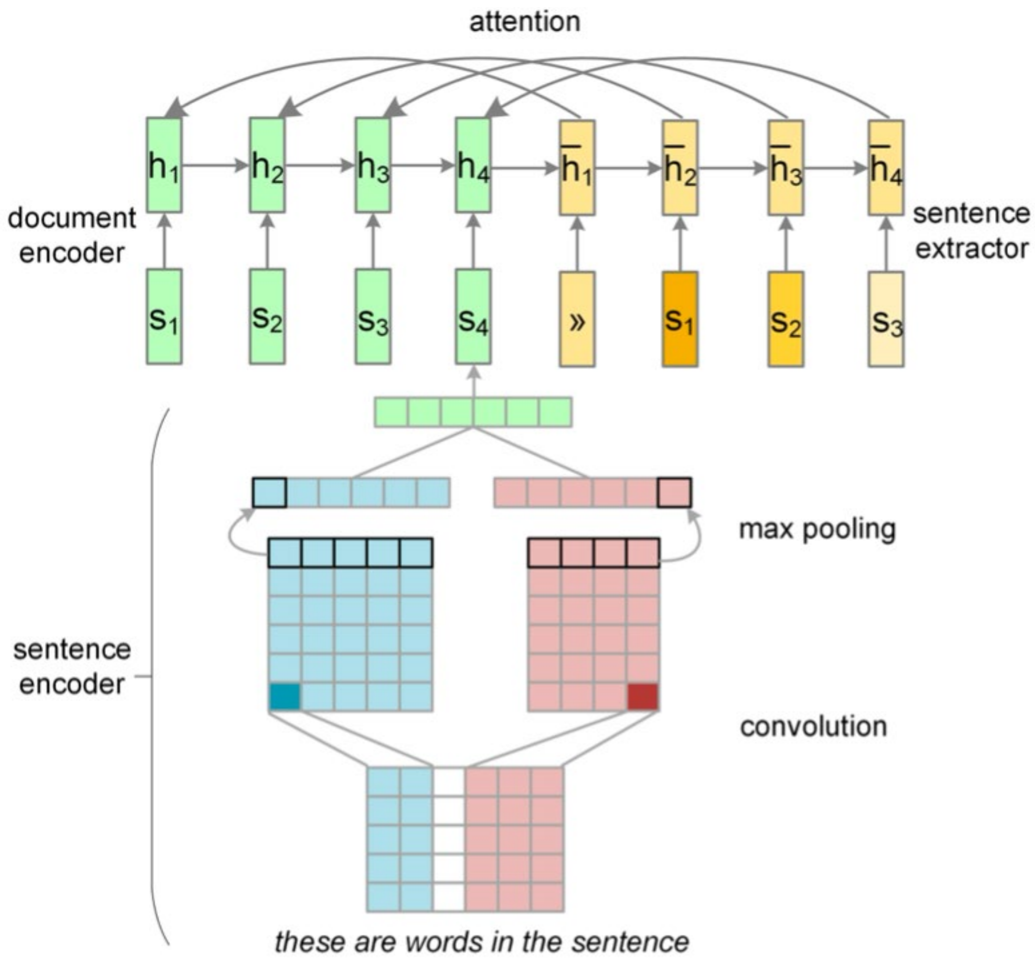
Gold Summary: Redpath has ended his eight-year association with Sale Sharks. Redpath spent five years as a player and three as a coach at sale. He has thanked the owners, coaches and players for their support.	Salience	Content	Novelty	Position	Prob.
Bryan Redpath has left his coaching role at Sale Sharks with immediate effect.	0.1	0.1	0.9	0.1	0.3
The 43 - year - old Scot ends an eight-year association with the Aviva Premiership side, having spent five years with them as a player and three as a coach.	0.9	0.6	0.9	0.9	0.7
Redpath returned to Sale in June 2012 as director of rugby after starting a coaching career at Gloucester and progressing to the top job at Kingsholm .	0.8	0.5	0.5	0.9	0.6
Redpath spent five years with Sale Sharks as a player and a further three as a coach but with Sale Sharks struggling four months into Redpath's tenure, he was removed from the director of rugby role at the Salford-based side and has since been operating as head coach .	0.8	0.9	0.7	0.8	**0.9**
'I would like to thank the owners, coaches, players and staff for all their help and support since I returned to the club in 2012.	0.4	0.1	0.1	0.7	0.2
Also to the supporters who have been great with me both as a player and as a coach,' Redpath said.	0.6	0.0	0.2	0.3	0.2



SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents

Extractive Summarization

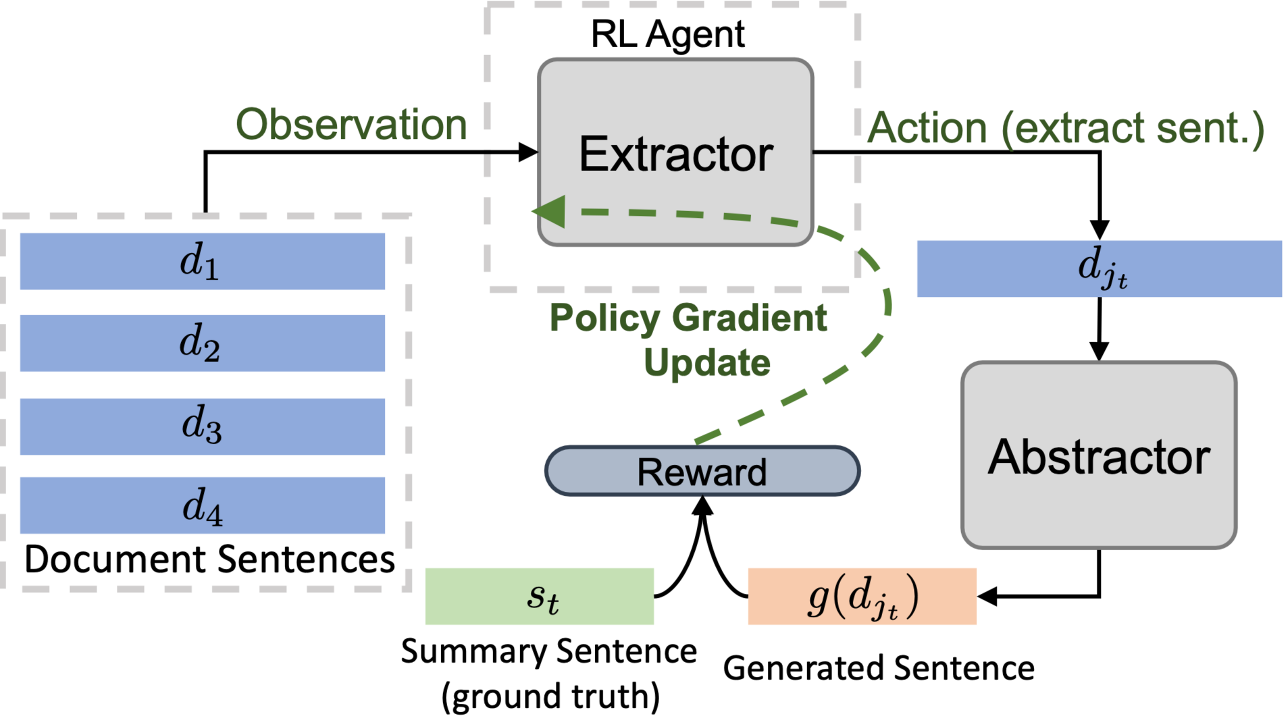
- Composed of a hierarchical document encoder and an attention-based extractor.
- Reader is to derive the meaning representation of a document based on its sentences and their constituent words
- Sentence extractor applies attention to directly extract salient sentences after reading them



Neural Summarization by Extracting Sentences and Words

Extractive Summarization

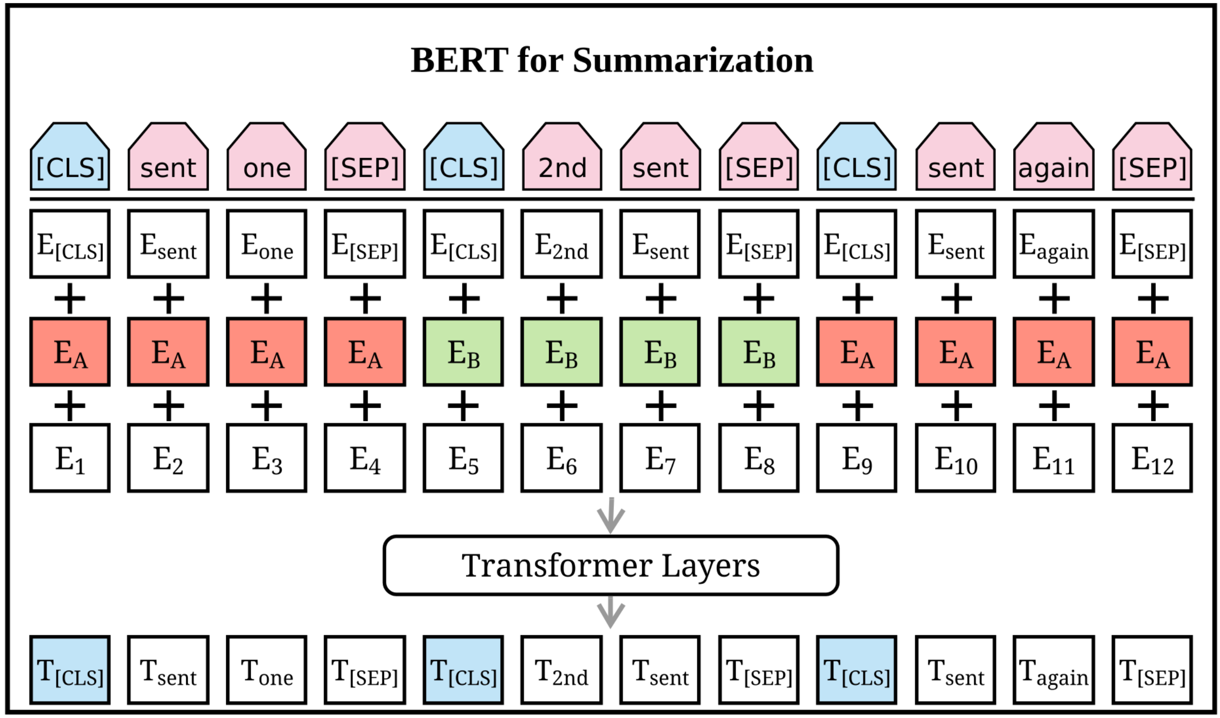
- Human usually select salient sentences and then rewrite them as the final summary.
- Sentence-level policy gradient method to bridge the non-differentiable computation between these two neural networks in a hierarchical way.



Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting

Extractive Summarization

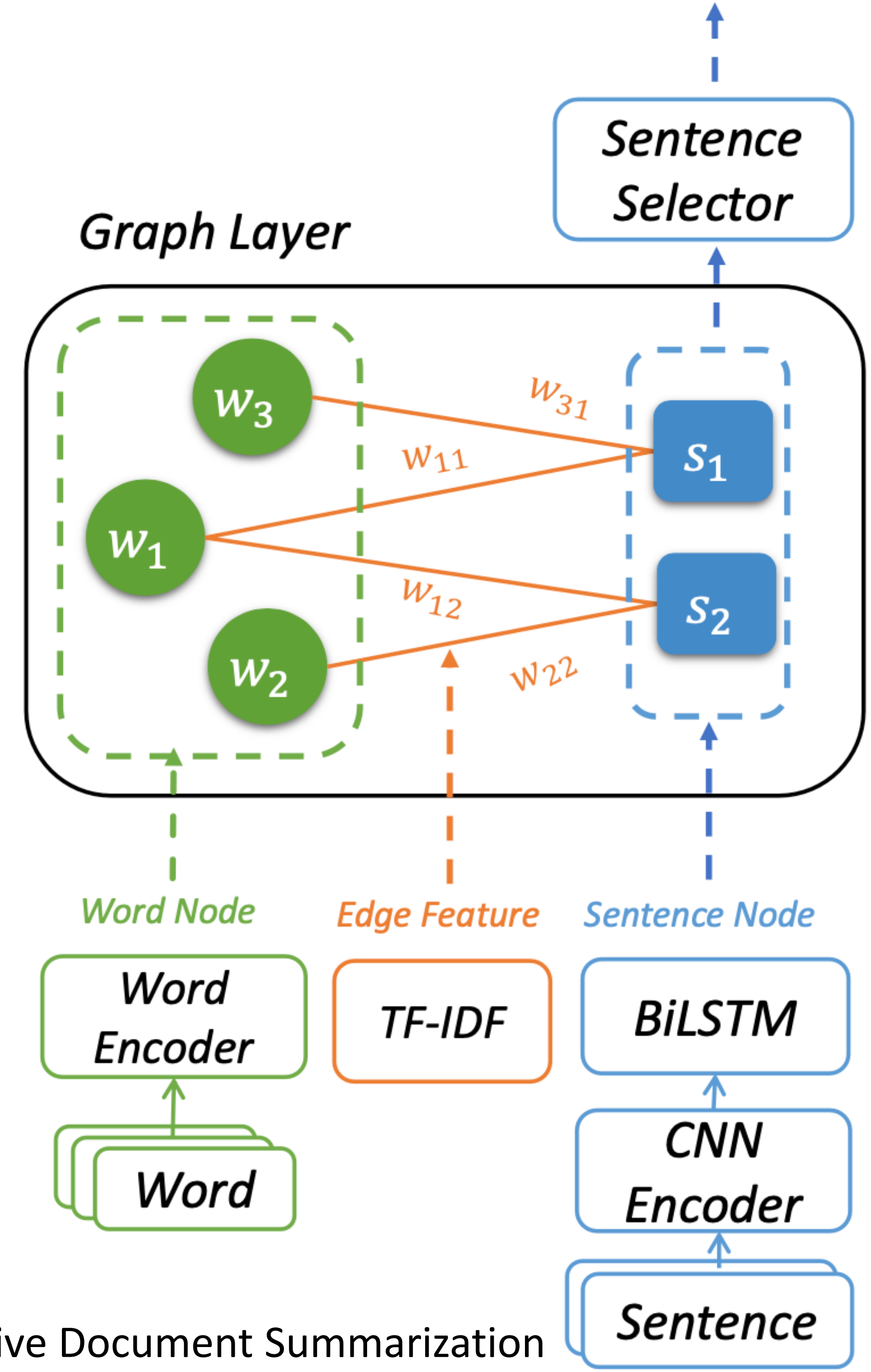
- Language model pretraining has advanced the state of the art in many NLP tasks
- Explore the potential of BERT for text summarization under a general framework
- Experiments on three datasets show that this model achieves state-of-the-art results



Text Summarization with Pretrained Encoders

Extractive Summarization

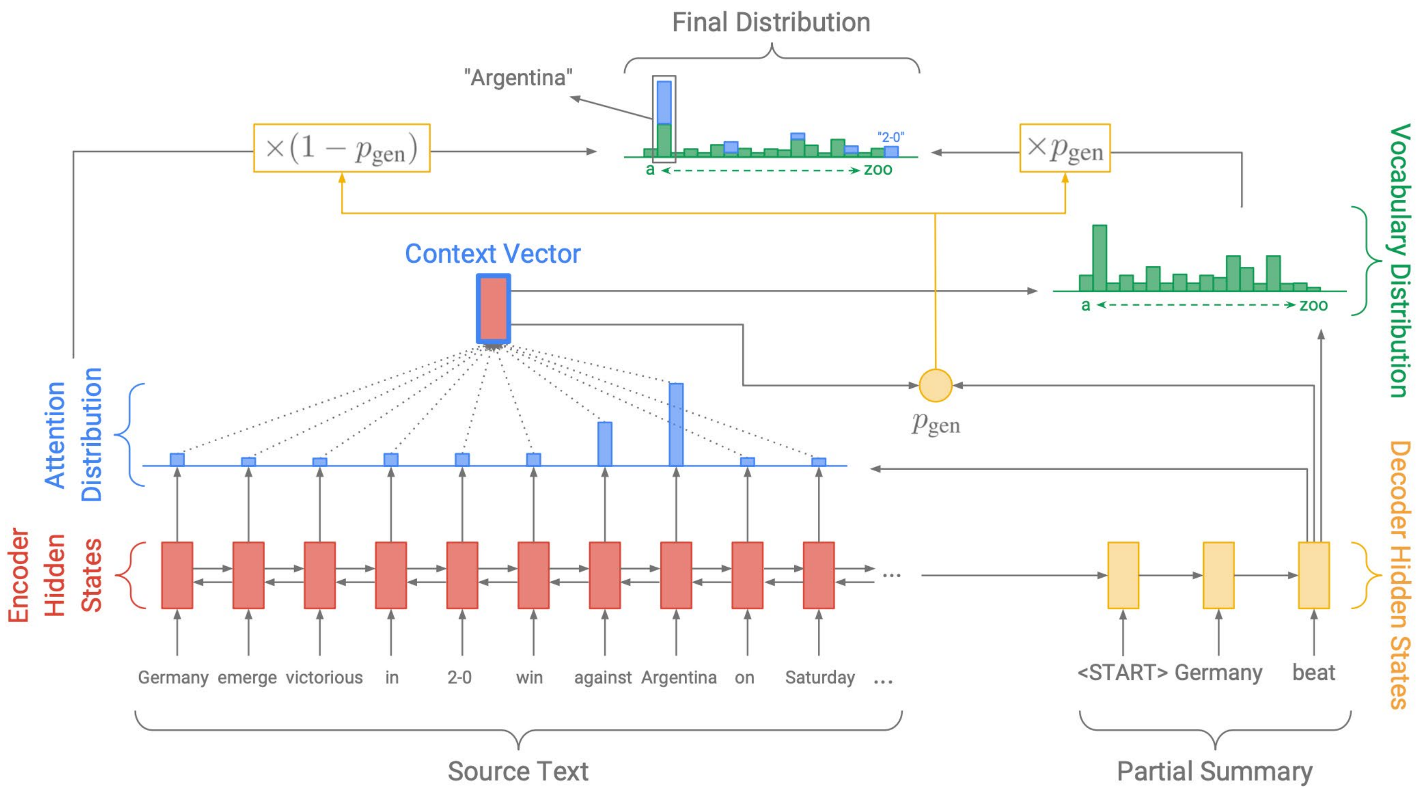
- Contains semantic nodes of different granularity levels apart from sentences
- These additional nodes act as the intermediary between sentences and enrich the cross-sentence relations.



Heterogeneous Graph Neural Networks for Extractive Document Summarization

Abstractive Summarization

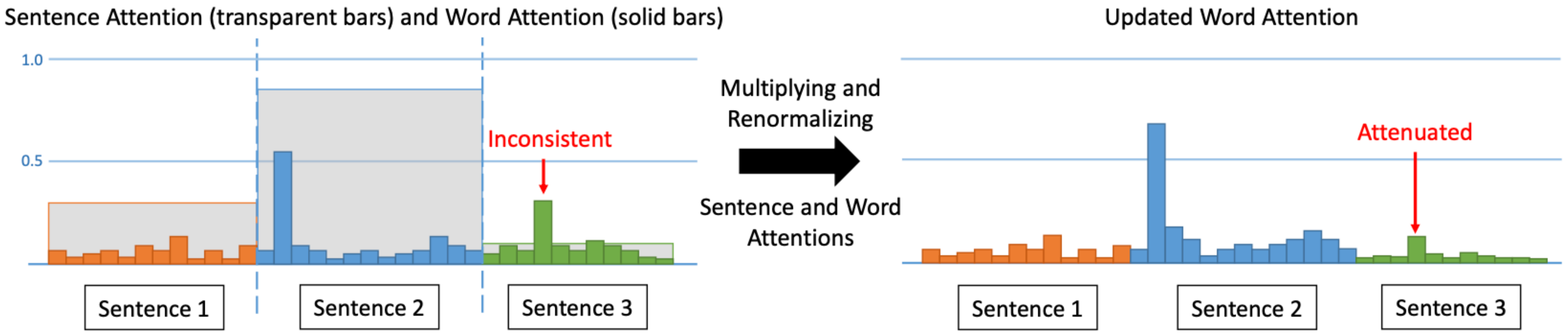
- Sequence-to-sequence models have provided a viable new approach for *abstractive* text summarization
- A hybrid pointer-generator network that can copy words from the source text via *pointing*, which aids accurate reproduction of information



Get To The Point: Summarization with Pointer-Generator Networks

Abstractive Summarization

- A unified model to combine the strength of both state-of-the-art extractor and abstracter.
- Inconsistency loss function is introduced to penalize the inconsistency between two levels of attentions.

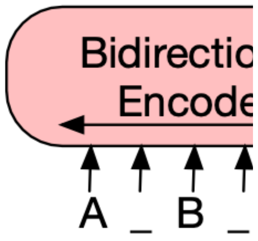


A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss

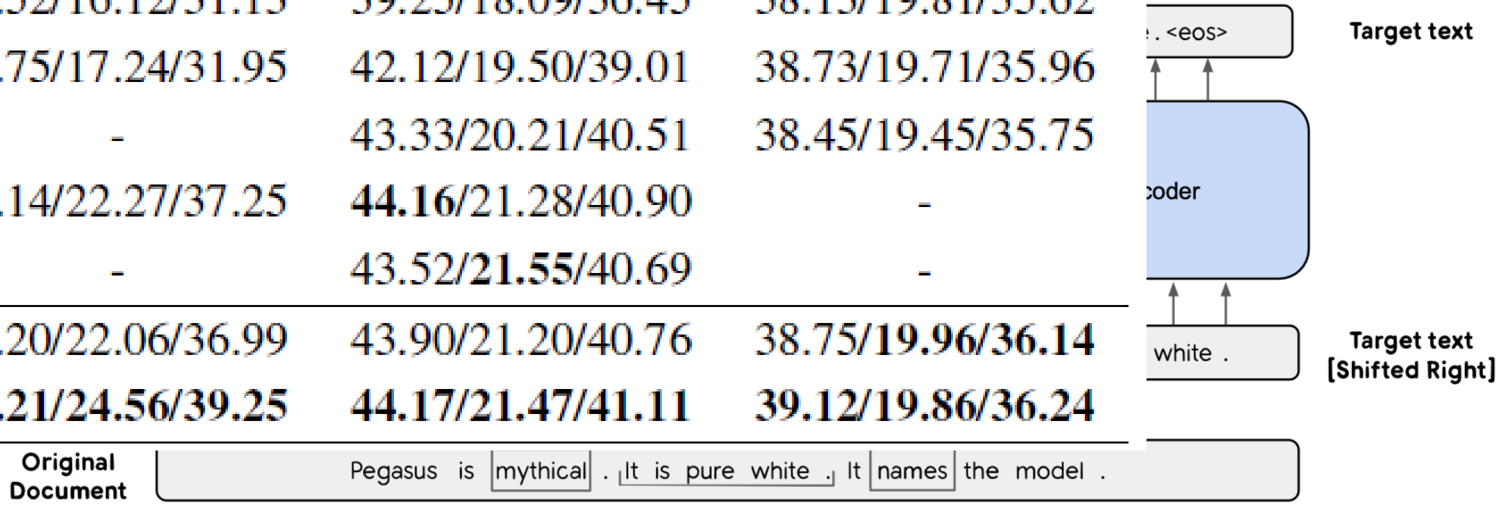
Abstractive Summarization

- Pre-training Transformers with self-supervised objectives on large text corpora has shown great success when fine-tuned on downstream NLP tasks including text summarization
- Important sentences are removed/masked from an input document and are generated together as one output sequence from the remaining sentences

R1/R2/RL	XSum	CNN/DailyMail	Gigaword
BERTShare (Rothe et al., 2019)	38.52/16.12/31.13	39.25/18.09/36.45	38.13/19.81/35.62
MASS (Song et al., 2019)	39.75/17.24/31.95	42.12/19.50/39.01	38.73/19.71/35.96
UniLM (Dong et al., 2019)	-	43.33/20.21/40.51	38.45/19.45/35.75
BART (Lewis et al., 2019)	45.14/22.27/37.25	**44.16**/21.28/40.90	-
T5 (Raffel et al., 2019)	-	43.52/**21.55**/40.69	-
PEGASUS$_{\text{LARGE}}$ (C4)	45.20/22.06/36.99	43.90/21.20/40.76	38.75/**19.96**/**36.14**
PEGASUS$_{\text{LARGE}}$ (HugeNews)	**47.21**/**24.56**/**39.25**	**44.17**/**21.47**/**41.11**	**39.12**/19.86/36.24



BART: Denoising Se
Language Generati



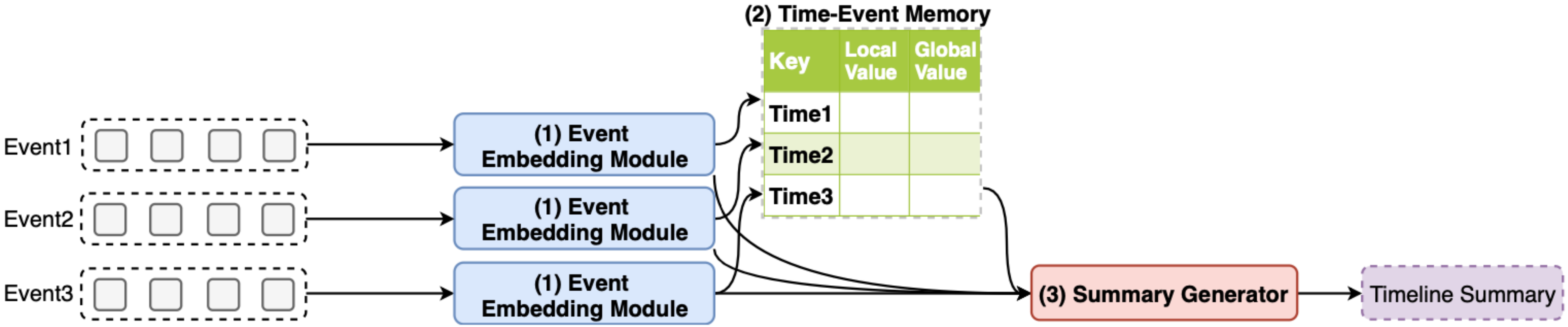
PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization

Incorporating Document Structure

- Timeline Summarization
  - help users to have a quick understanding of the overall evolution of any given topic
  - consider evolutionary characteristics of news plus to traditional summary elements
- Extreme Long Document Summarization
  - the input document can be very long, such as an academic paper or a patent document which is longer than the news article
  - extract the salient information and central idea from a large amount of information.
- Dialog Summarization
  - time-consuming for people to review all the context before starting a new dialog
  - the salient information is scattered in the whole dialog history
- Academic paper summarization
  - The reference relationship should be considered into generating summary of academic paper.
- Movie Summarization
  - Summarizing longer narratives, screenplays, whose form and structure is far removed from newspaper articles.

Timeline Summarization

- Timeline summarization is an important research task which can help users to have a quick understanding of the overall evolution of any given topic.
- The previous works are all based on extractive methods
- A large-scale dataset with 169,423 training samples, 5,000 evaluation and 5,000 test samples.
- On average, there are 352.22 words and 61.16 words in article and summary respectively.



Learning towards Abstractive Timeline Summarization

Timeline Summarization

- Given any collection of time-stamped news articles, MTLS automatically discovers important yet different stories and generates a corresponding timeline for each story.
- Propose a Two-Stage Affinity Propagation Summarization framework which is a two-stage clustering-based framework.

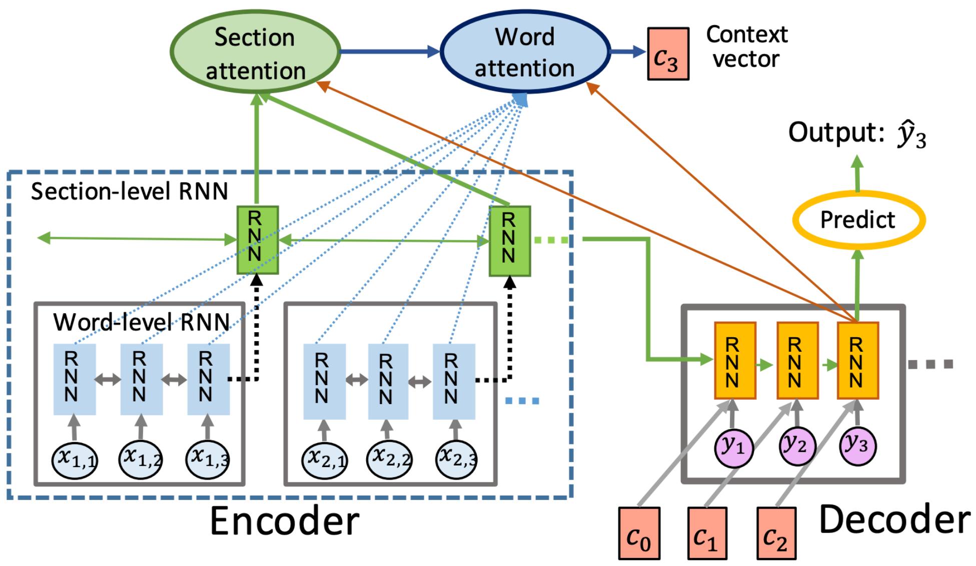
MTLS Methods		concat		align+m:1		agreement		d-select
		ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2	F1
Baselines								
CHIEU2004	Random	0.191	0.027	0.019	0.004	0.010	0.002	0.075
	LDA	0.192	0.035	0.023	0.005	0.013	0.004	0.089
	k-means	0.229	0.046	0.027	0.006	0.014	0.004	0.096
MARTSCHAT2018	Random	0.254	0.049	0.044	0.009	0.037	0.007	0.352
	LDA	0.289	0.068	0.062	0.017	0.052	0.015	0.387
	k-means	0.291	0.071	0.061	0.017	0.051	0.015	0.376
GHALANDARI2020	Random	0.253	0.048	0.068	0.015	0.058	0.013	0.414
	LDA	0.268	0.062	0.085	0.025	0.076	0.024	0.440
	k-means	0.284	0.073	**0.096**	0.030	0.085	0.028	0.467
Our method								
2SAPS		**0.312**	**0.084**	**0.096**	**0.033**	**0.089**	**0.029**	**0.556**

Multi-TimeLine Summarization (MTLS): Improving Timeline Summarization by Generating Multiple Summaries

Extreme Long Document Summarization

- Datasets of long document summarization task
- A hierarchical encoder, capturing the discourse structure of the document.
- A discourse-aware decoder that generates the summary.

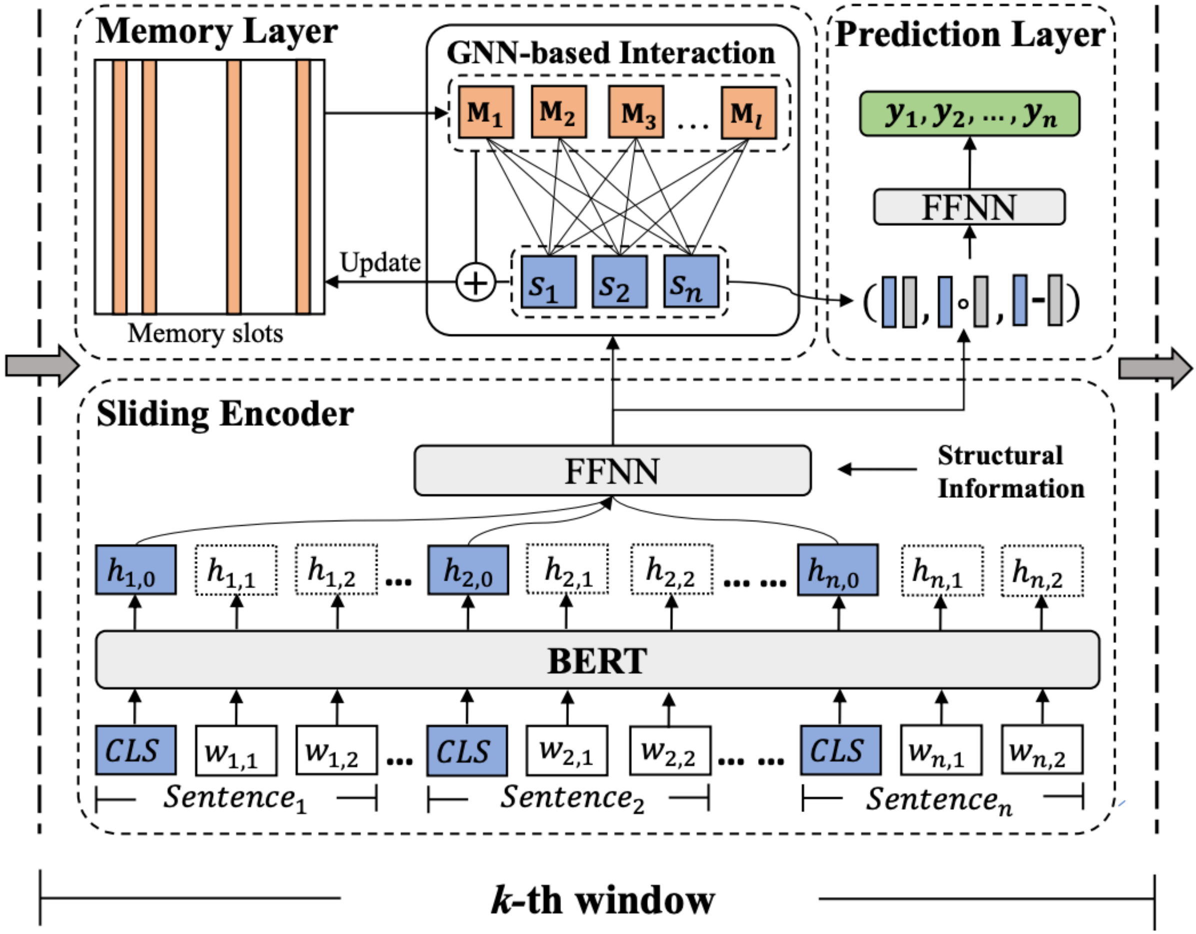
Dataset	# Doc	Summary # word	Summary # sent	Doc # word	Comp.	Den.
PUBMED	133,215	202.4	6.8	3049.0	16.2	5.8
ARXIV	215,913	272.7	9.6	6029.9	39.8	3.8
BILLSUM	23,455	207.7	7.2	1813.0	13.6	4.1
BIGPATENT	1,341,362	116.5	3.7	3573.2	36.3	2.4
GOVREPORT	19,466	**553.4**	**17.8**	**9409.4**	19.0	7.3



A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents

Extreme Long Document Summarization

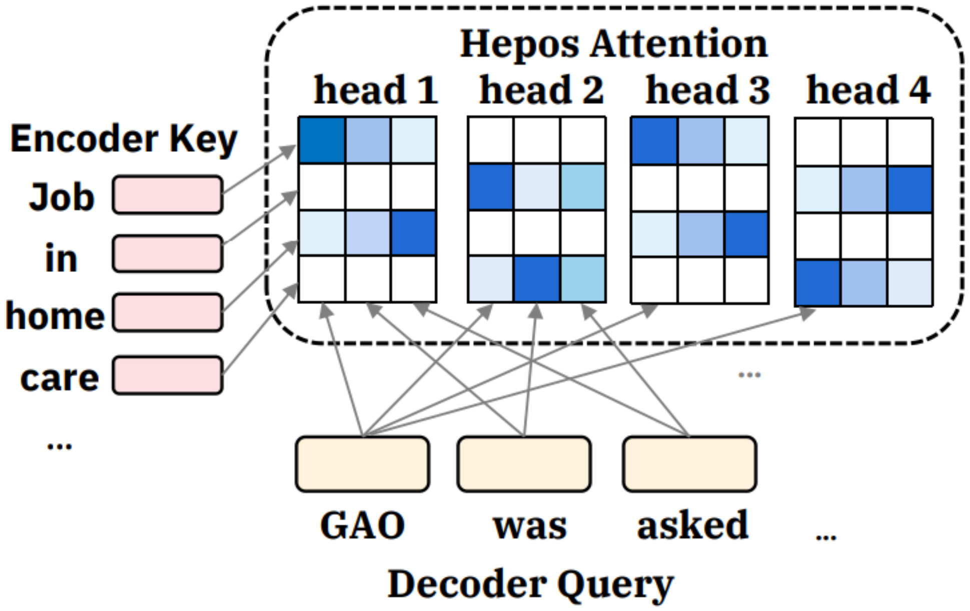
- Sliding selector network with dynamic memory for extractive summarization of long-form documents
- A memory to preserve salient information learned from previous windows



Sliding Selector Network with Dynamic Memory for Extractive Summarization of Long Documents

Extreme Long Document Summarization

- The main challenge of summarizing long document is how to find salient information from large amount of sentences effectively.
- Encoder-decoder attention with head-wise positional strides



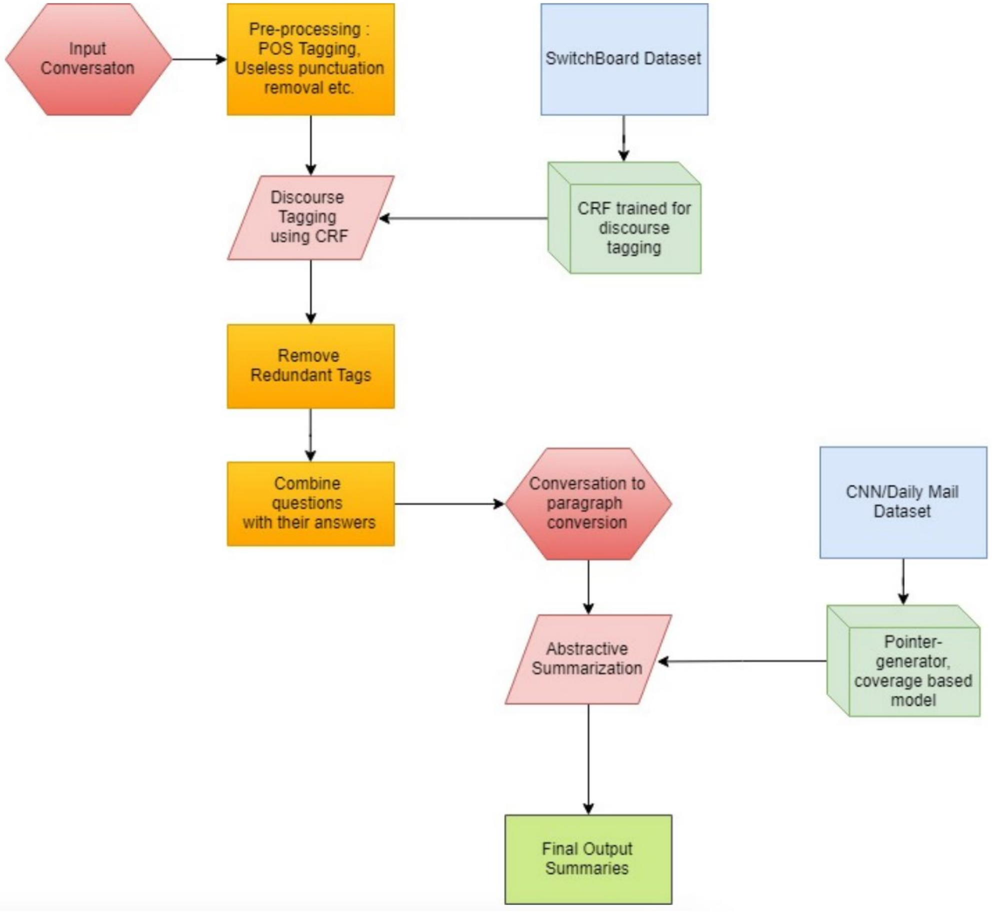
Dialog Summarization - Datasets

Dataset	MEDIASUM	AMI	ICSI	DiDi	CRD3	MultiWOZ	SAMSum
Source	Transcribed Speech					Written	
Type	Interview	Meeting	Meeting	Customer	Game	Booking	Daily
Real dialogue	✓	✓	✓	✓	✓	✓	×
Open domain	✓	×	×	×	×	×	✓
Public	✓	✓	✓	×	✓	✓	✓
Dialogues	463,596	137	59	328,880	159	10,438	16,369
Dial. words	1,553.7	4,757	10,189	/	31,802.8	180.7	83.9
Summ. words	14.4	322	534	/	2062.3	91.9	20.3
Turns	30.0	289	464	/	2,507.4	13.7	9.9
Speakers	6.5	4	6.2	2	9.6	2	2.2

MEDIASUM: A Large-scale Media Interview Dataset for Dialogue Summarization

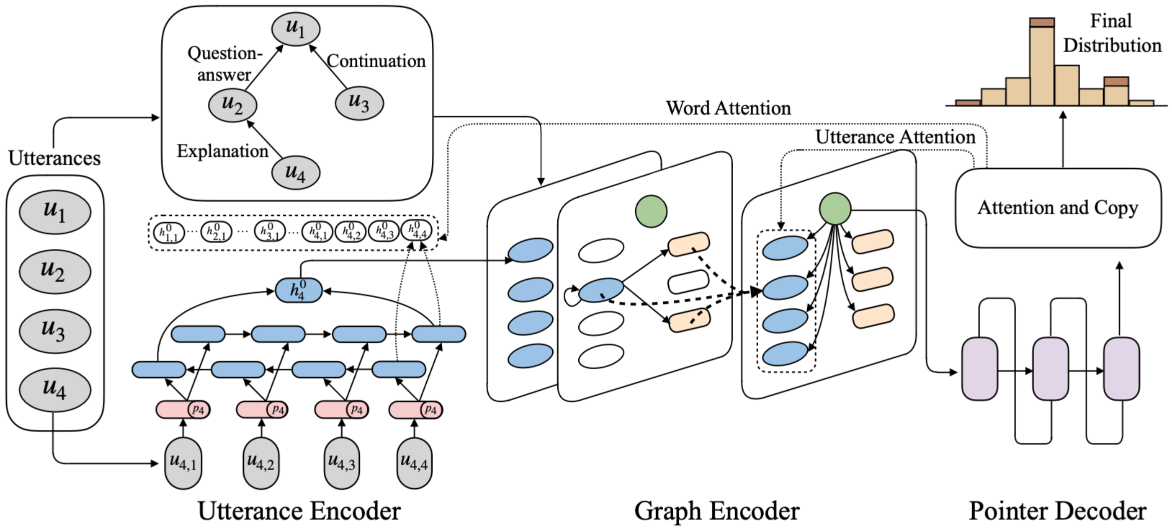
Dialog Summarization

- In this section we describe the complete pipeline of the model which includes (1) Sequence labelling of utterance tags , (2) Re-ordering of conversation to model discourse relations, and (3) Pointer-generator, coverage based model for abstractive summarization.



Dialog Summarization

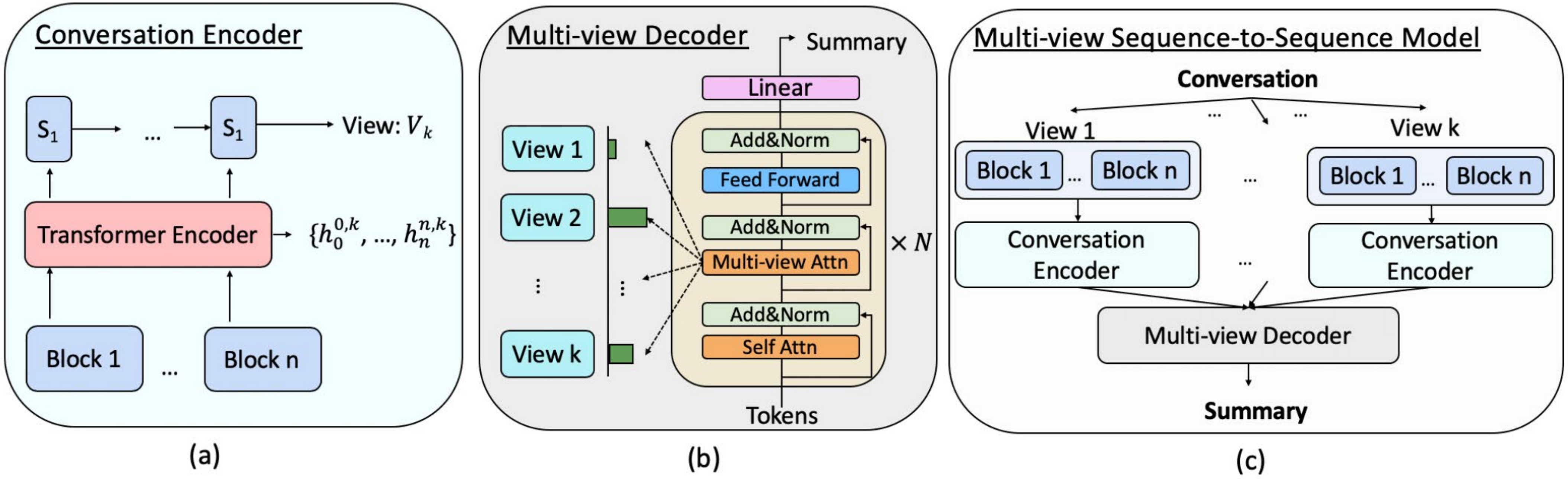
- A meeting is naturally full of dialogue-specific structural information
- Previous works model a meeting in a sequential manner, while ignoring the rich structural information
- Dialogue discourse is a dialogue-specific structure that can provide pre-defined semantic relationships between each utterance



Dialogue Discourse-Aware Graph Convolutional Networks for Abstractive Meeting Summarization

Dialog Summarization

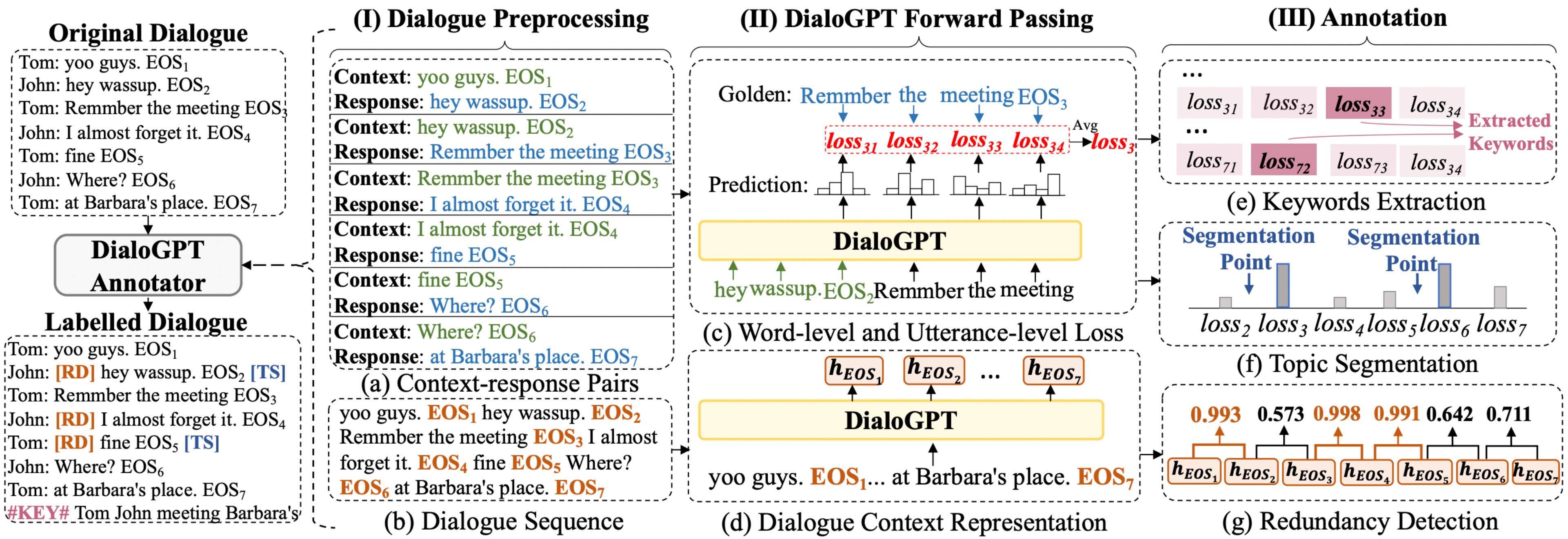
- **Topic View**: Based on what topics were discussed,it can be segmented into *several topics*
- **Stage view**: From a conversation progression perspective
- **Global View:** Conversations can be treated as a whole
- **Discreate View**: Each utterance can serve as one segment



Multi-View Sequence-to-Sequence Models with Conversational Structure for Abstractive Dialogue Summarization

Dialog Summarization

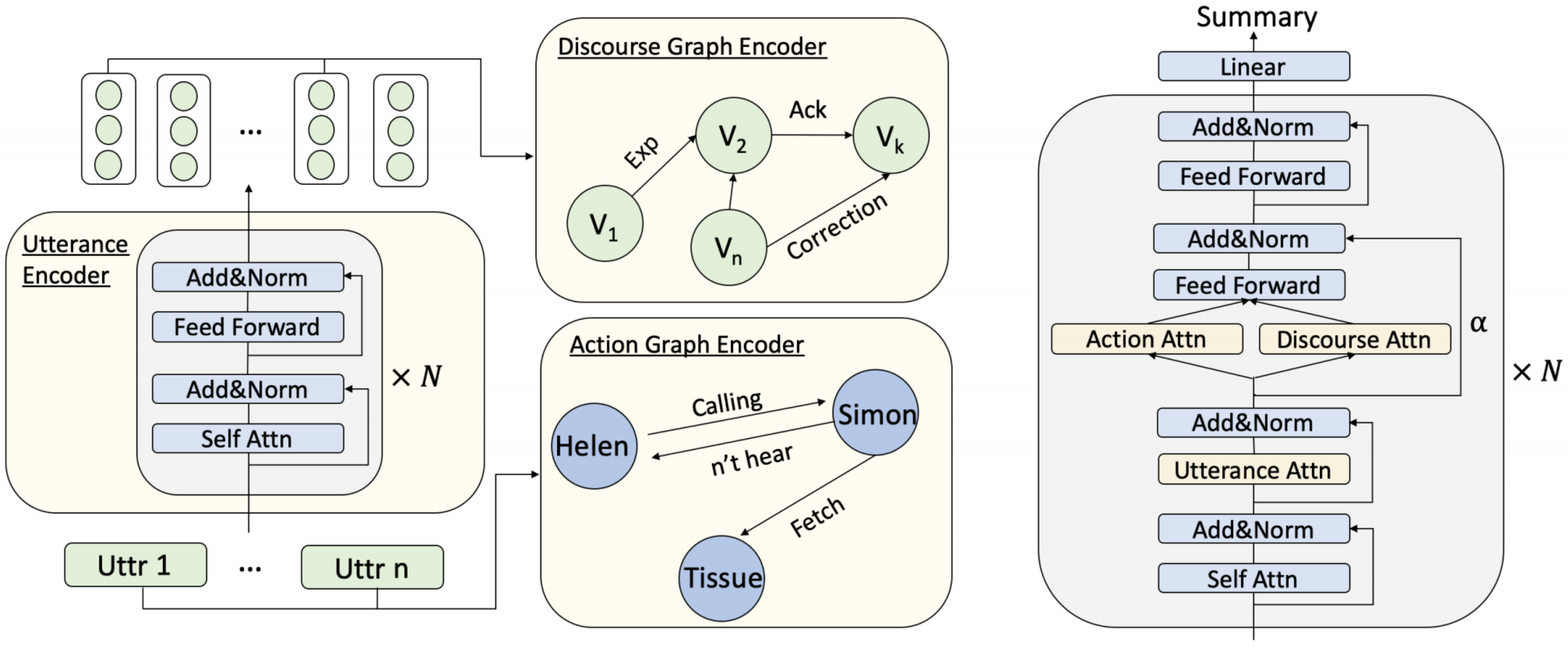
- Existing features are obtained via open-domain toolkits that are dialog-agnostic or heavily relied on human annotations
- Perform three dialogue annotation tasks takes advantage of dialogue background knowledge encoded in DialoGPT



Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization

Dialog Summarization

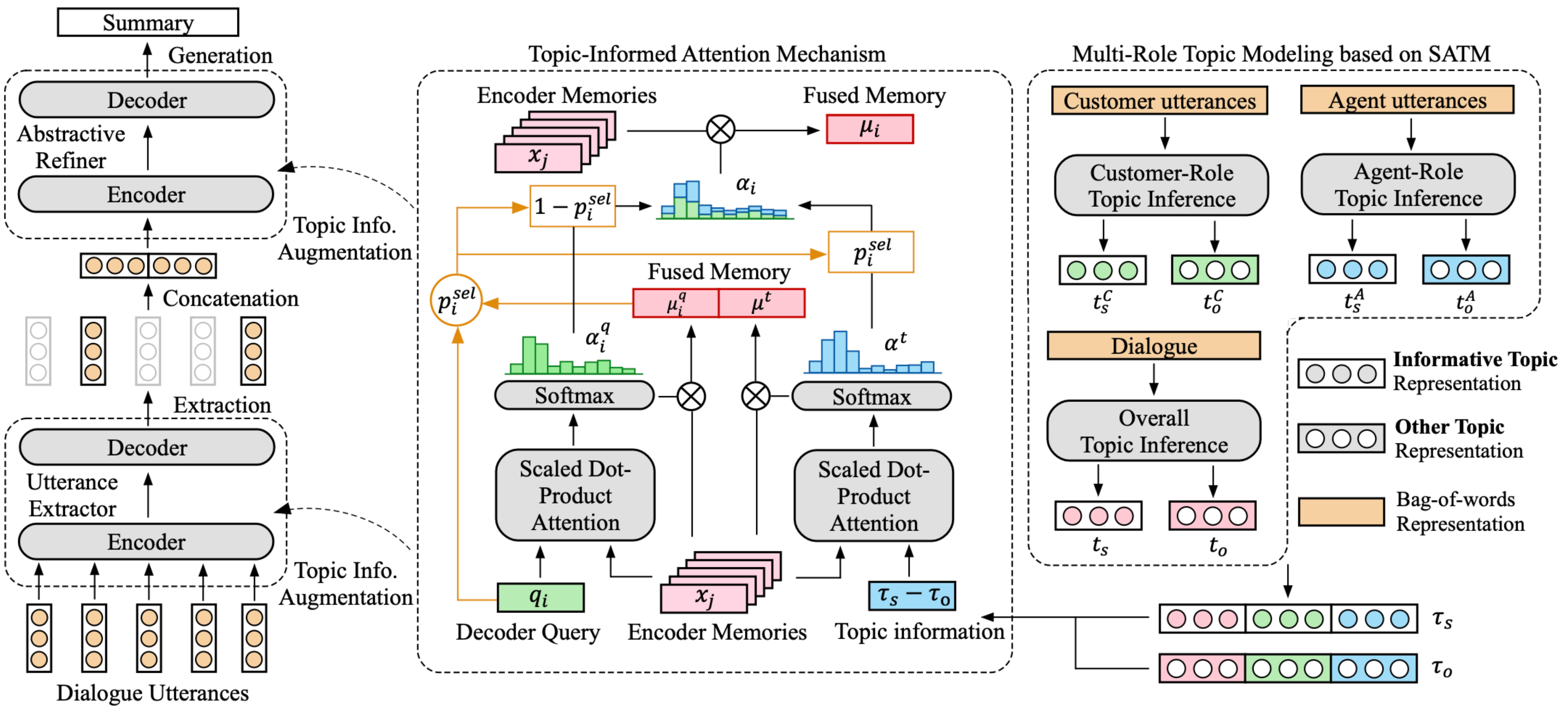
- Existing generated dialog summaries often suffer from insufficient, redundant, or incorrect content
- Explicitly model the rich structures in conversations for more precise and accurate conversation summarization



(a) Utterance and Graph Encoder (b) Multi-granularity Decoder

Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs

Dialog Summarization



Topic-Oriented Spoken Dialogue Summarization for Customer Service with Saliency-Aware Topic Modeling

Academic paper summarization

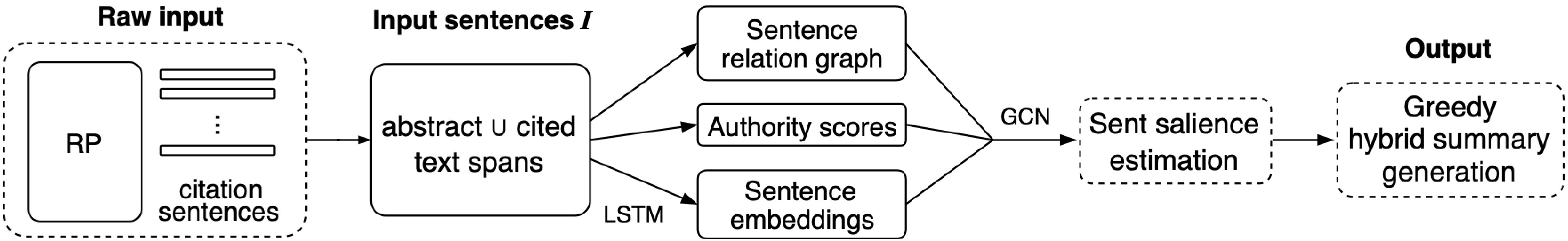
- Datasets

Datasets	Source	# Pairs Train	# Pairs Val	# Pairs Test	Doc. Length # Words	Doc. Length # Sent.	Sum. Length # Words	Sum. Length # Sent.	# Sections
CNN	News	90,266	1,220	1,093	760.5	34.0	45.7	3.6	-
DailyMail	News	196,961	12,148	10,397	653.3	29.3	54.7	3.9	-
ScisummNet	Scientific Papers	1009	–	–	4203.4	178.0	150.7	7.4	6.5
arXiv†	Scientific Papers	215,913	6440	6436	4938.0	206.3	220.0	9.6	5.9
PubMed†	Scientific Papers	119,924	6633	6658	3016.0	86.4	203.0	6.9	5.6
SSN (inductive)	Scientific Papers	128,400	6123	6276	5072.3	290.6	165.1	6.4	10.8
SSN (transductive)		128,299	6250	6250					

Enhancing Scientific Papers Summarization with Citation Graph

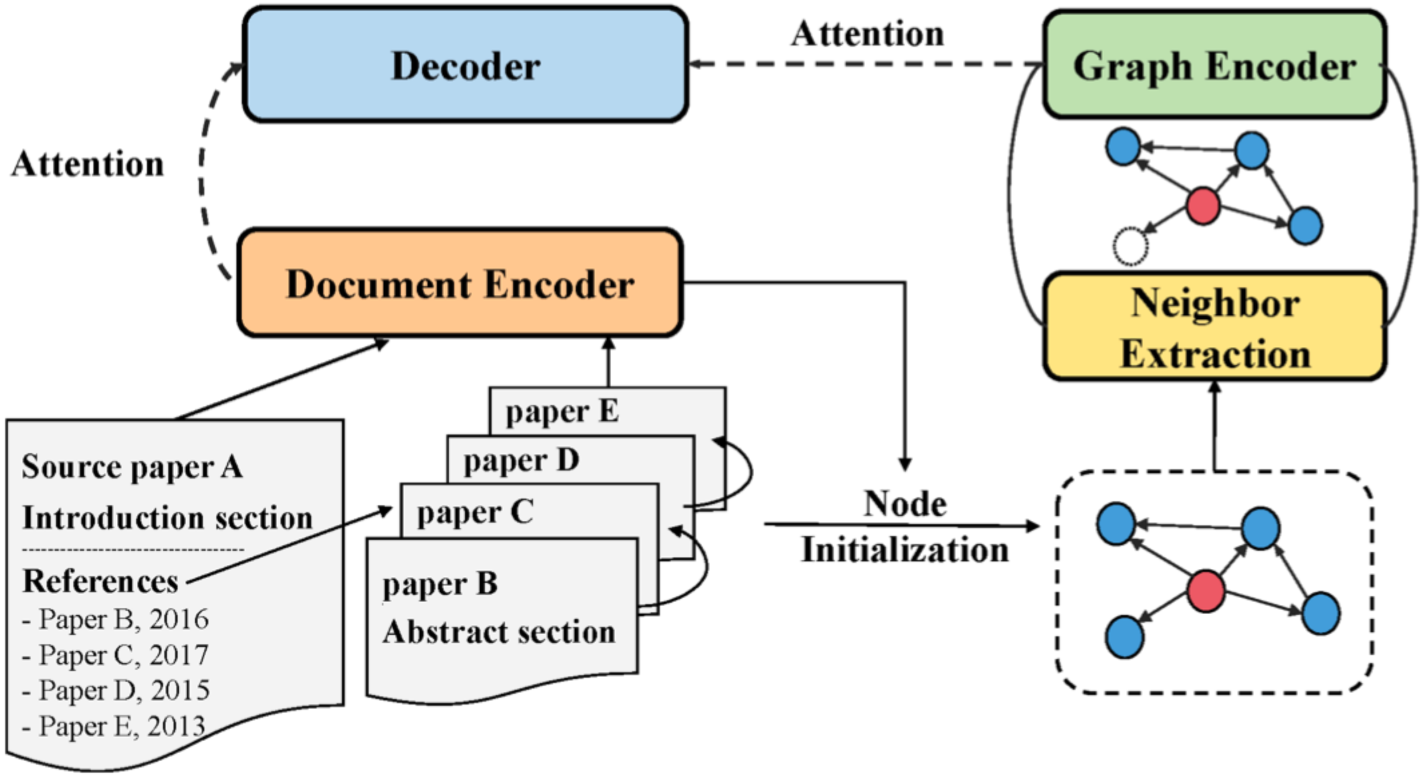
Academic paper summarization

- Integrating the authors' original highlights (abstract) and the article's actual impacts on the community



ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks

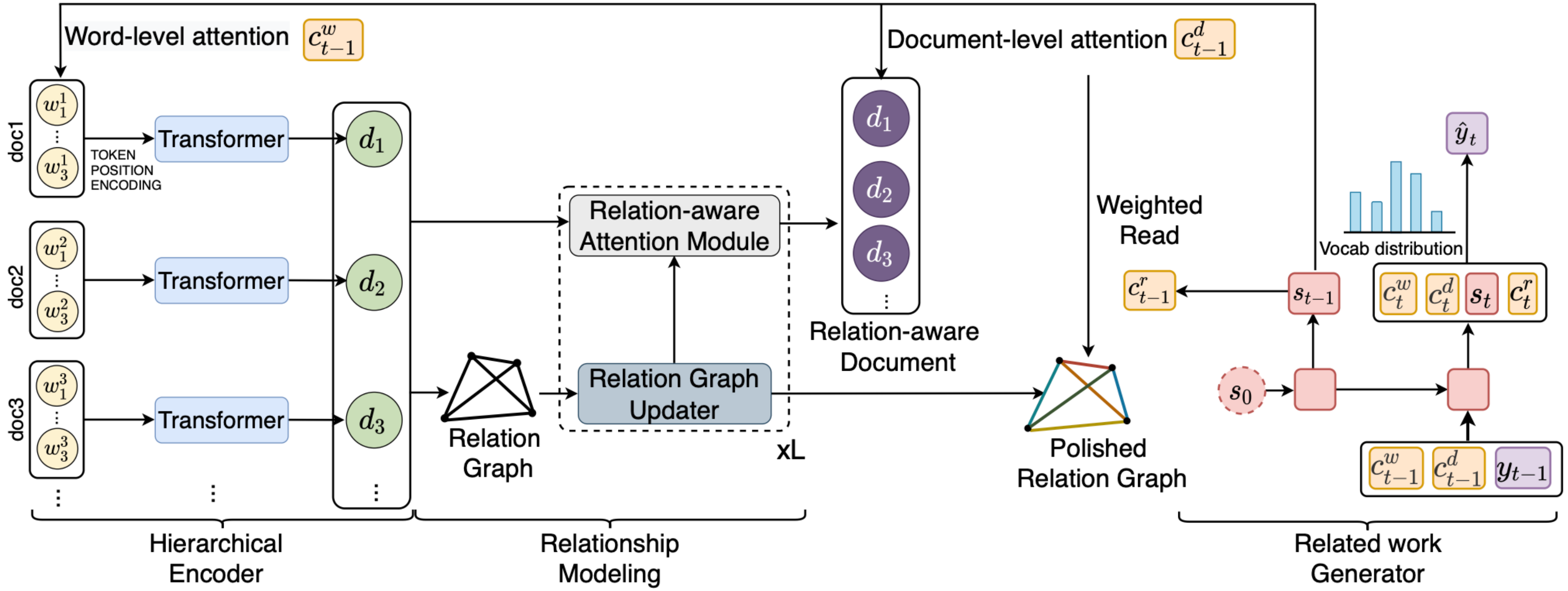
Academic paper summarization



Enhancing Scientific Papers Summarization with Citation Graph

Academic paper summarization

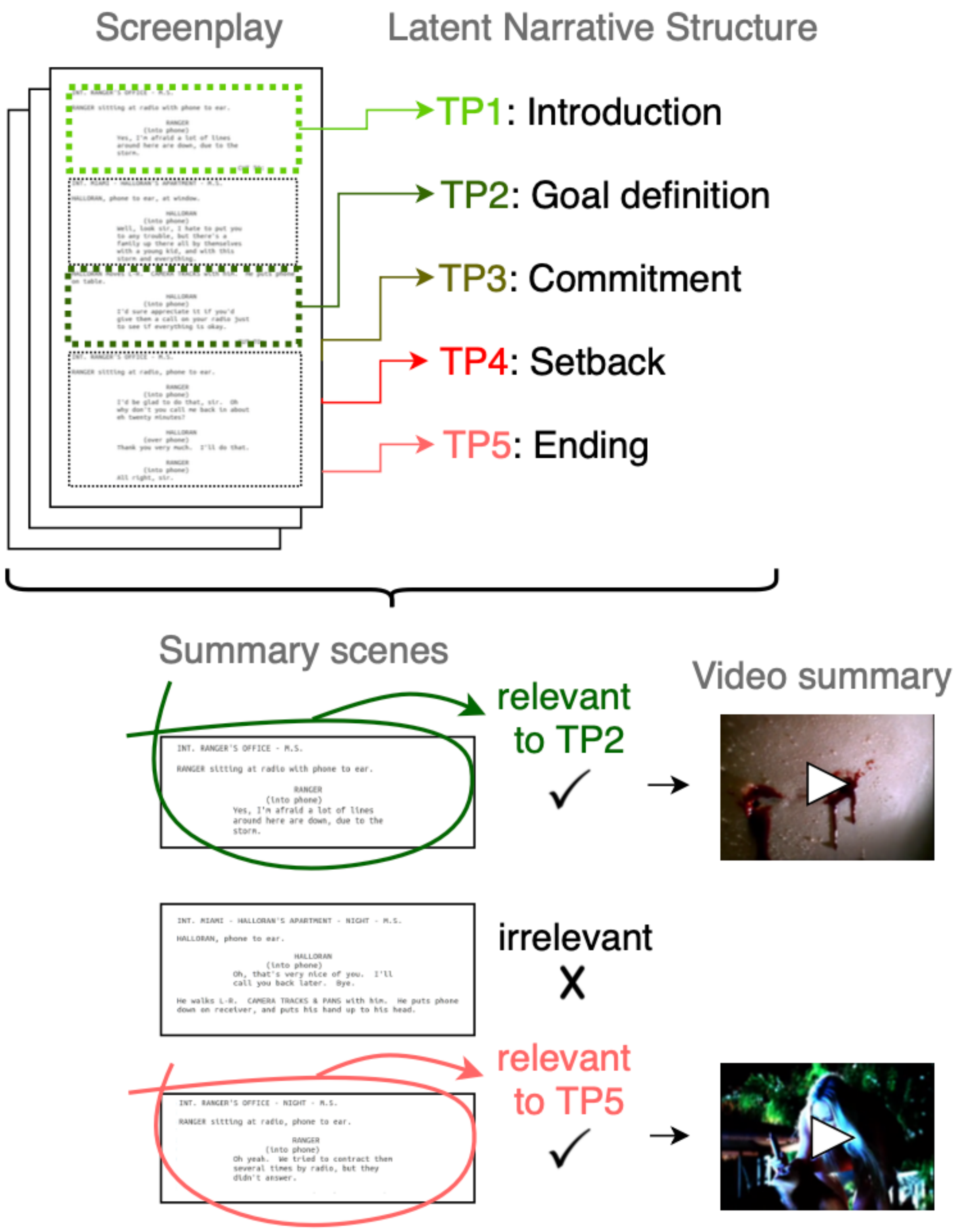
- Given a set of related publications, related work section generation aims to provide researchers with an overview of the specific research area by summarizing these works and introducing them in a logical order.



Capturing Relations between Scientific Papers: An Abstractive Model for Related Work Section Generation

Movie Summarization

- Most efforts to date have concentrated on the summarization of news articles
- Screenplays, whose form and structure is far removed from newspaper articles.



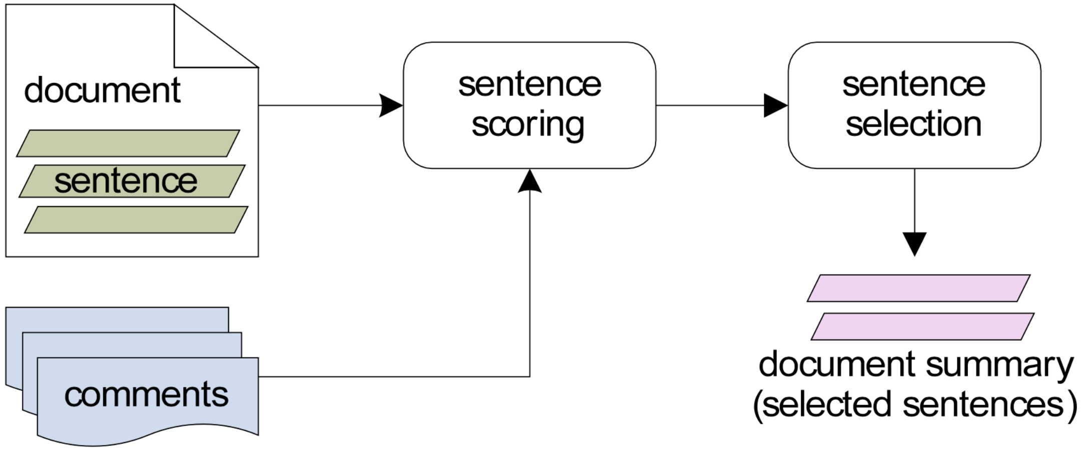
Screenplay Summarization Using Latent Narrative Structure

Incorporating Additional Knowledge

- Reader-aware Summarization
  - reader comments concentrate on the main idea of the news article
  - comments can be used to help the summarization model to capture the main idea
- Template Based Summarization
  - first retrieves a summary template and then edits it into the new summary of the current document.
- Multi-Modal Summarization
  - increase of multi-media data on the web
  - the visual information is incorporated along with the input document into the text summarizing process to improve the quality
- Query-based Summarization
  - search engine provides a list of web pages associated with their summaries
  - should summarize the query focused aspect of the web page instead of the main idea
- Opinion Summarization
  - Summarize the opinion of e-commerce reviews

Reader-aware Summarization

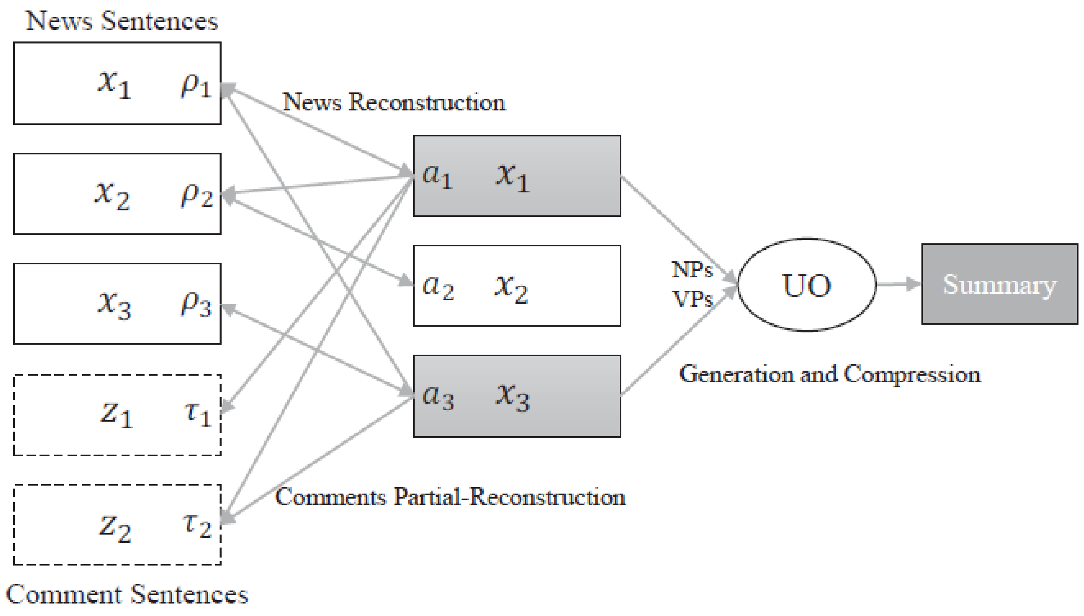
- In the beginning, researchers firstly propose to understand the input document with the feedback of readers using a graph-based method, where they identify three relations (topic, quotation, and mention) by which comments can be linked to one another.



Comments-Oriented Document Summarization: Understanding Documents with Readers' Feedback

Reader-aware Summarization

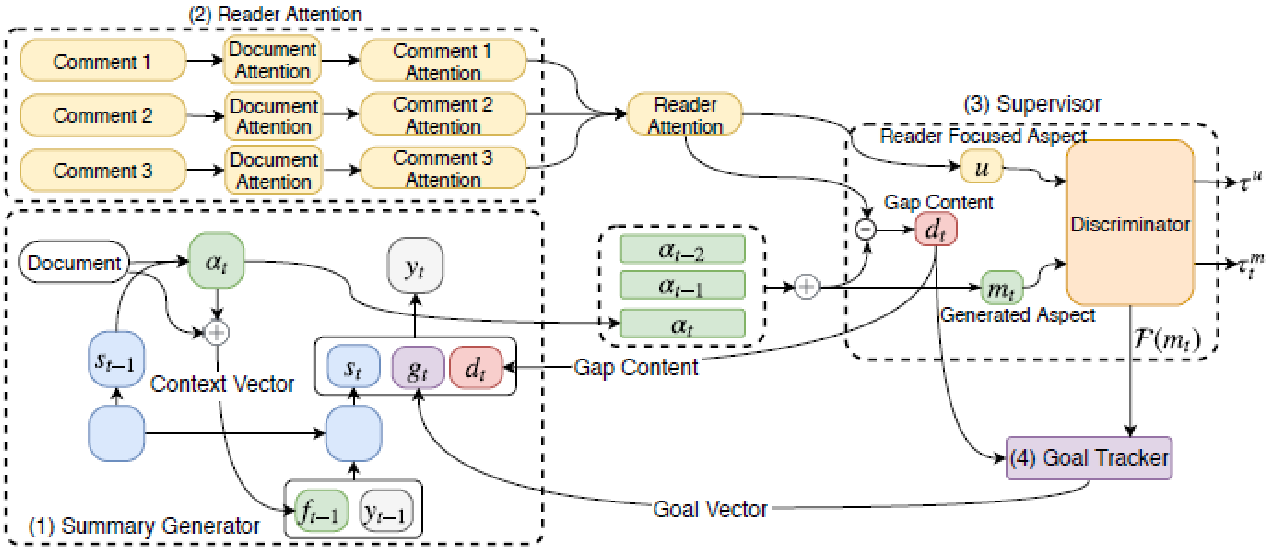
Employ a sparse coding based framework for this task which jointly considers news documents and reader comments via an unsupervised data reconstruction strategy.



Reader-Aware Multi-Document Summarization via Sparse Coding

Reader-aware Summarization

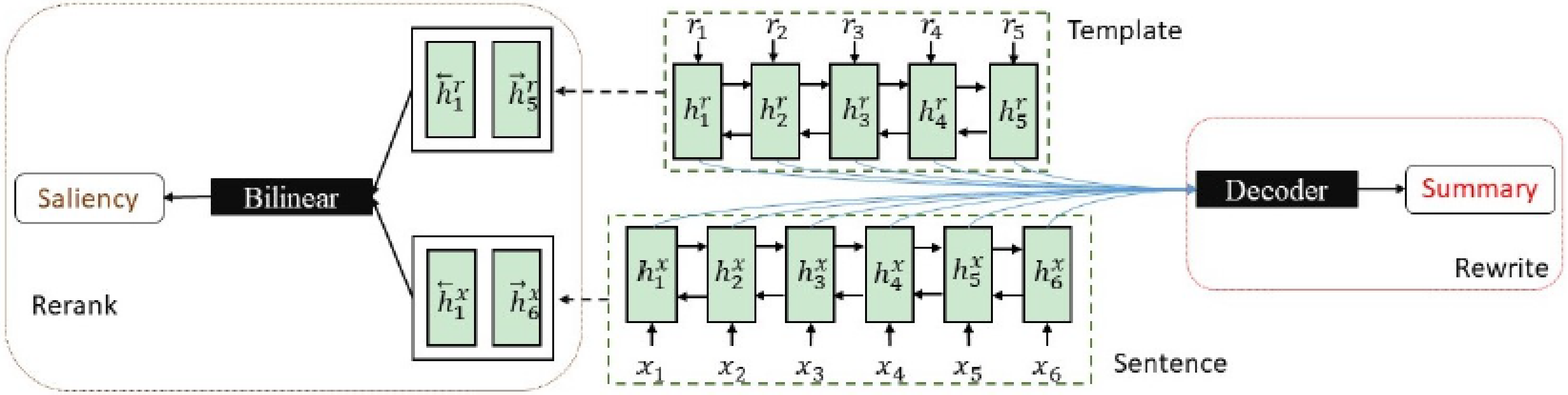
- A large-scale reader-aware summarization dataset (863826 training samples)
- A generative-adversarial based method which conducts the interaction between reader comments and news to capture the reader attention distribution on the article



Abstractive Text Summarization by Incorporating Reader Comments

Template Based Summarization

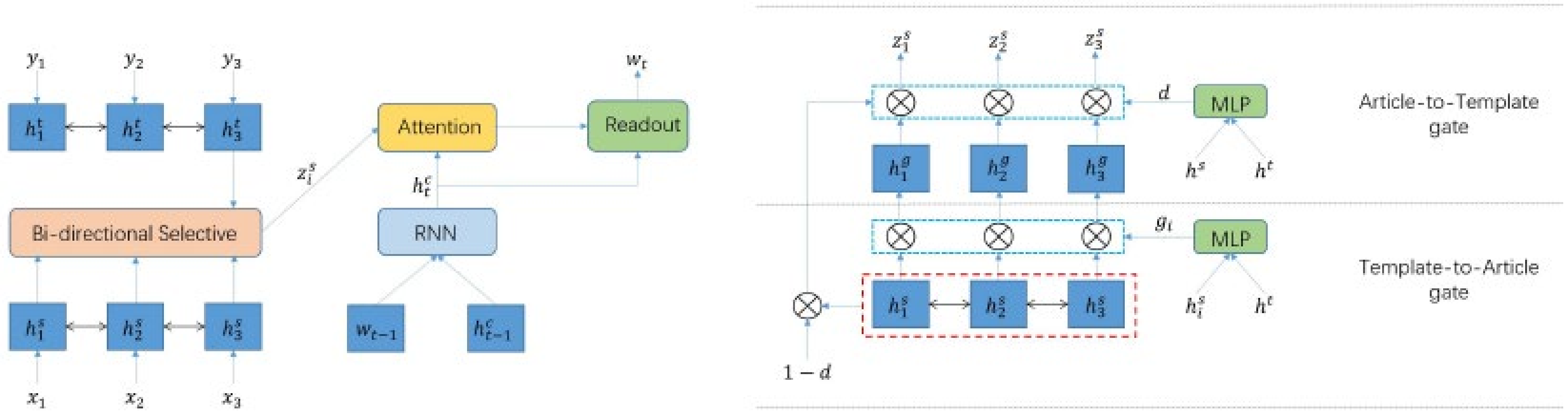
- Previous seq2seq purely depend on the source, which tends to work unstably
- Use a popular IR platform to Retrieve proper summaries as candidate templates
- Extend the seq2seq framework to jointly conduct template Reranking and template-aware summary generation



Retrieve, Rerank and Rewrite: Soft Template Based Neural Summarization

Template Based Summarization

- Bi-directional Selective Encoding with Template (BiSET) model
- Leverages template discovered from training data to softly select key information from each source article
- A multi-stage process for automatic retrieval of high-quality templates from training corpus.



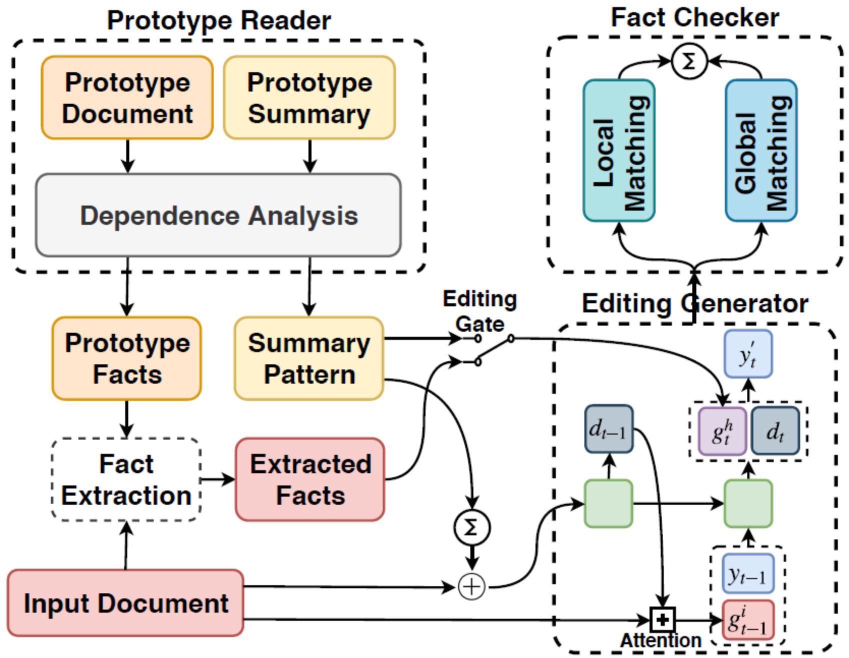
(a)

(b)

BiSET: Bi-directional Selective Encoding with Template for Abstractive Summarization

Template Based Summarization

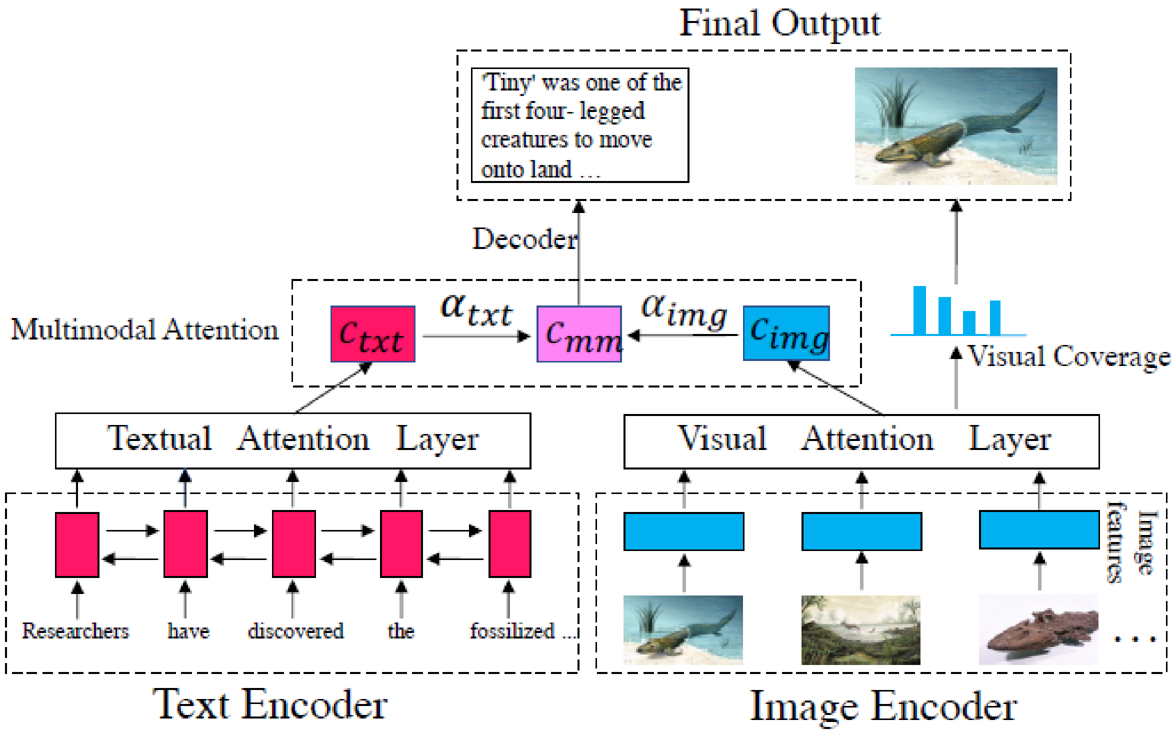
- In circumstances, the generated summaries are required to conform to a specific pattern
- Template-based methods are too rigid for our patternized summary generation task.
- We propose a summarization framework na med Prototype Editing based Summary Gen erator that incorporates prototype documen t-summary pairs



How to Write Summaries with Patterns? Learning towards Abstractive Summarization through Prototype Editing

Multi-Modal Summarization - Image

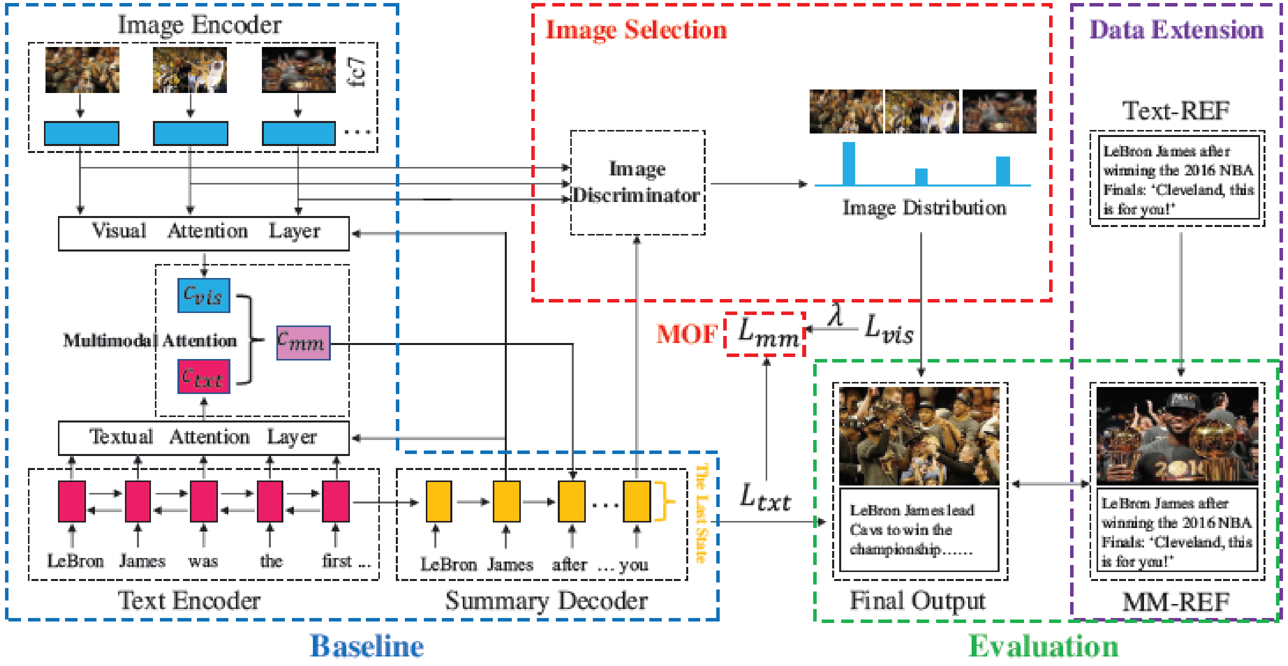
- Multimodal Summarization with Multimodal Output
- Four modules: text encoder, image encoder, multimodal attention layer, and summary decoder
- Propose a multimodal automatic evaluation (MMAE) method which mainly consid- ers three aspects: salience of text, salience of image, and relevance between text and image.



MSMO: Multimodal Summarization with Multimodal Output

Multi-Modal Summarization - Image

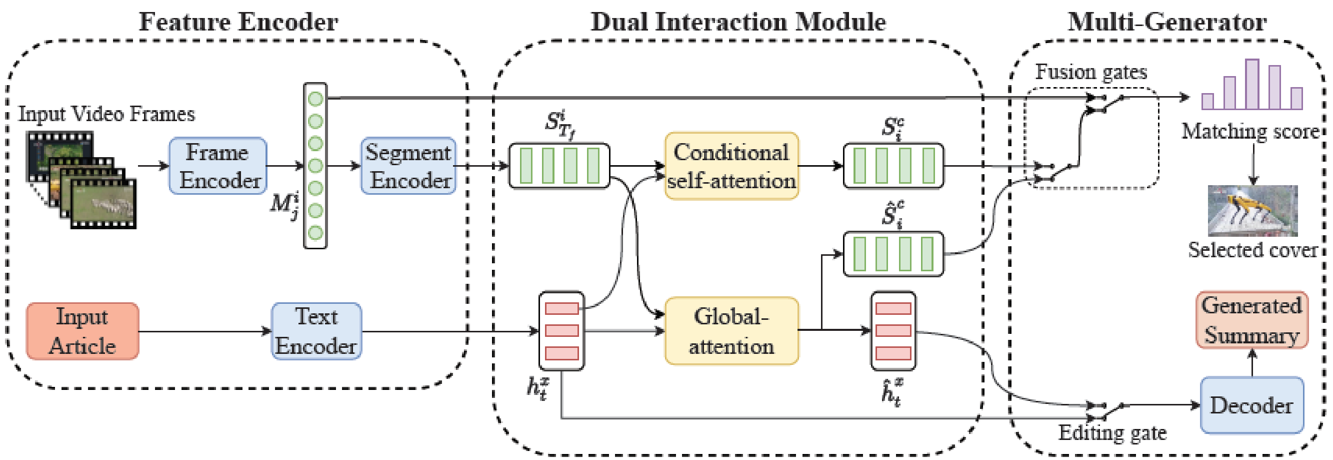
- Existing MSMO methods are trained by the target of text modality
- Leading to the modality-bias problem
- Propose a multimodal objective function with the guidance of multimodal reference



Multimodal Summarization with Guidance of Multimodal Reference

Multi-Modal Summarization - Video

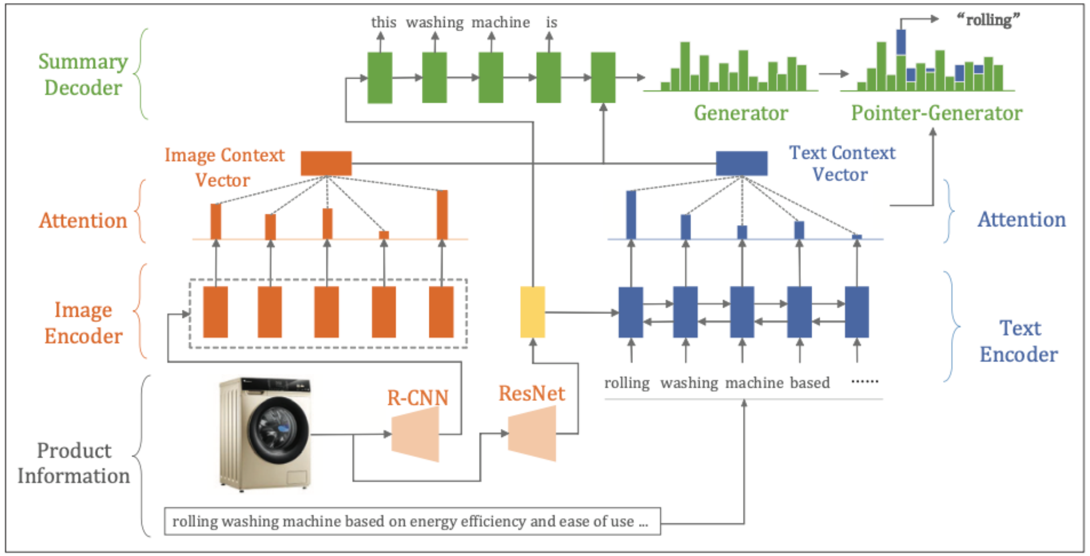
- Video and document as input
- Selects cover frame from news video and generates textual summary of the news article in the meantime



VMSMO: Learning to Generate Multimodal Summary for Video-based News Articles

Multi-Modal Summarization – Ecommerce

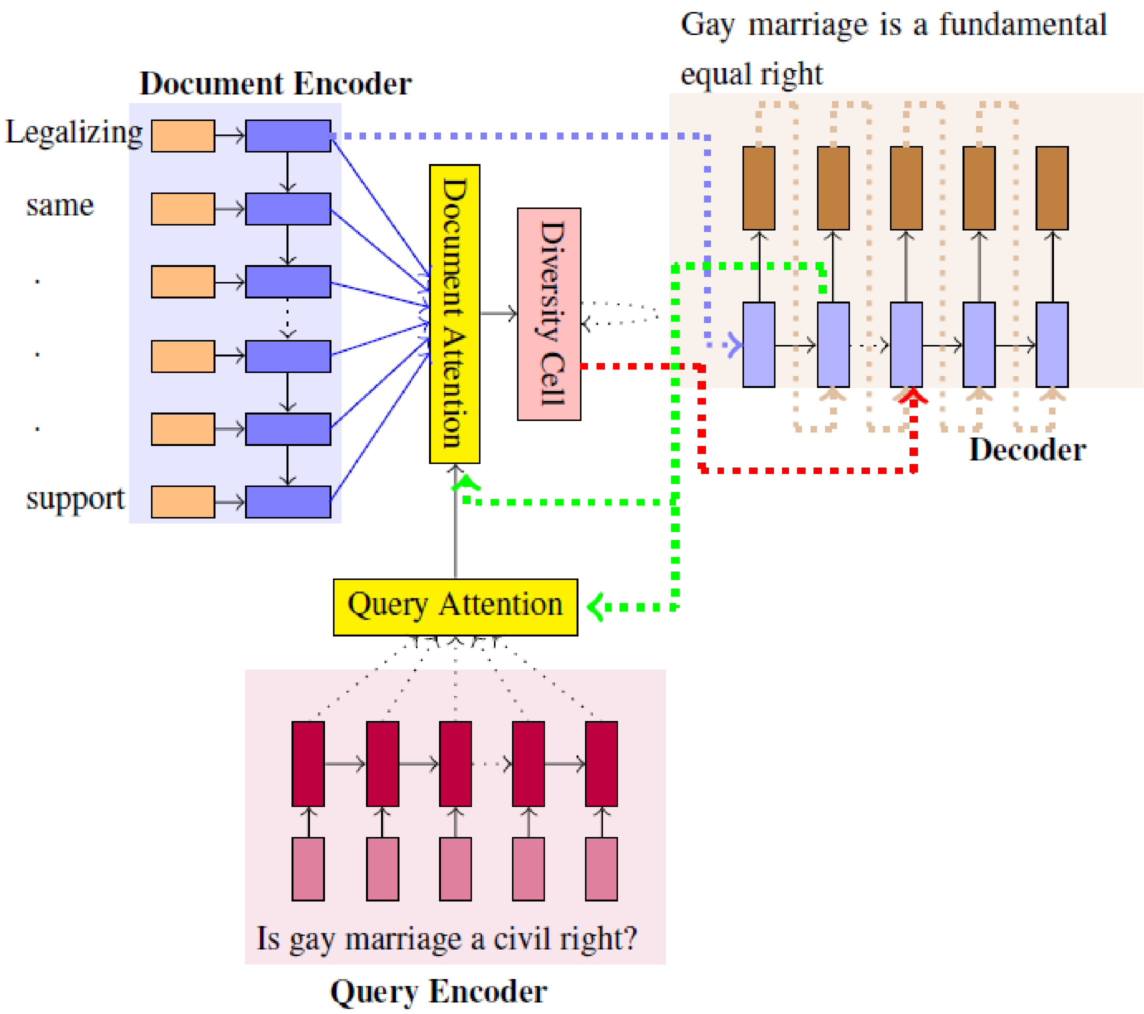
- Adopt an aspect-based reward augmented maximum likelihood training method
- Aspect coverage mechanism to keep track of what aspects have been mentioned
- Adopt constrained decoding to enhance the coherence of summaries



Aspect-Aware Multimodal Summarization for Chinese E-Commerce Products

Query-based Summarization

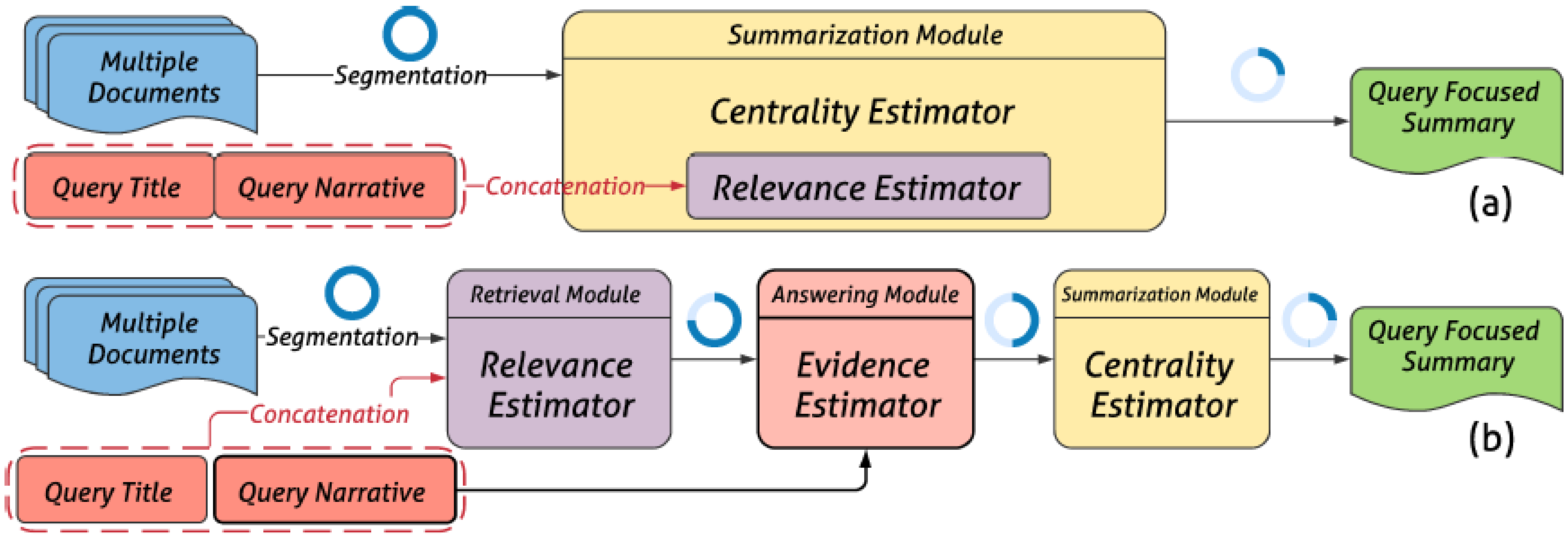
- Query-based summarization highlights those points that are relevant to the user query
- Seq2seq suffers from the drawback of generation of repeated phrases
- A query attention model which learns to focus on different portions of the query
- A new diversity based attention model



Diversity driven attention model for query-based abstractive summarization

Query-based Summarization

- A coarse-to-fine modeling framework for extractive query focused summarization which incorporates a *relevance* estimator, an *evidence* estimator and a *centrality* estimator.



Coarse-to-Fine Query Focused Multi-Document Summarization

Query-based Summarization

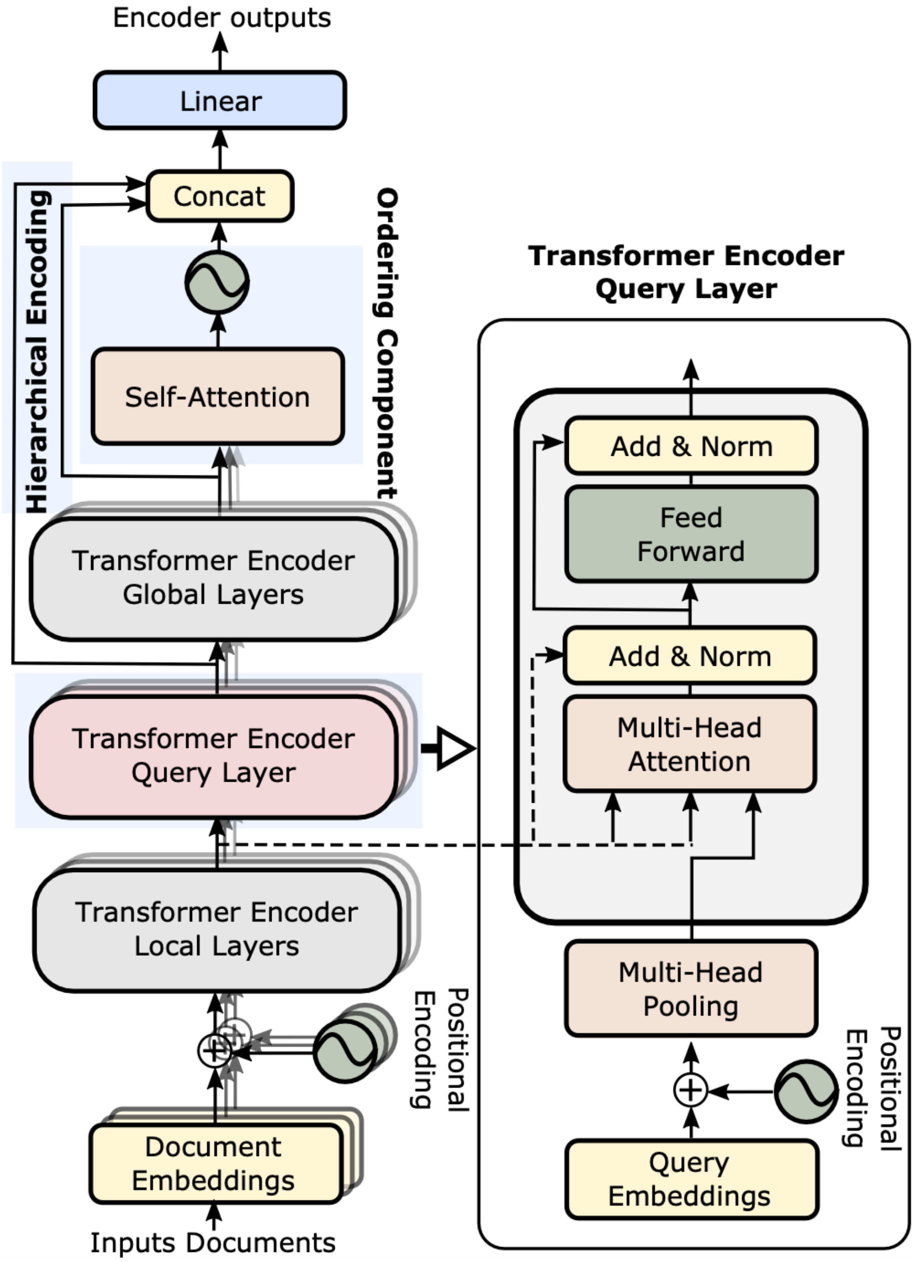
- Existing methods are limited by the lack of sufficient large-scale high-quality training datasets.
- Present two QMDS training datasets: (1) QMDSCNN and (2) QMDSIR by using two data augmentation methods

Statistics	Train	Val	Test
QMDSCNN (# samples)	287,113	13,368	11,490
- Avg. # documents	6.5	6.5	6.5
- Avg. Doc. length (# tokens)	355	346	353
- Avg. Query length (# tokens)	13.8	14.5	14.2
QMDSIR (# samples)	82,076	10,259	10,260
- Avg. # documents	5.8	5.4	5.5
- Avg. Doc. length (# tokens)	1,291	1,402	1,379
- Avg. Query length (# tokens)	6.2	6.2	6.2

Data Augmentation for Abstractive Query-Focused Multi-Document Summarization

Query-based Summarization

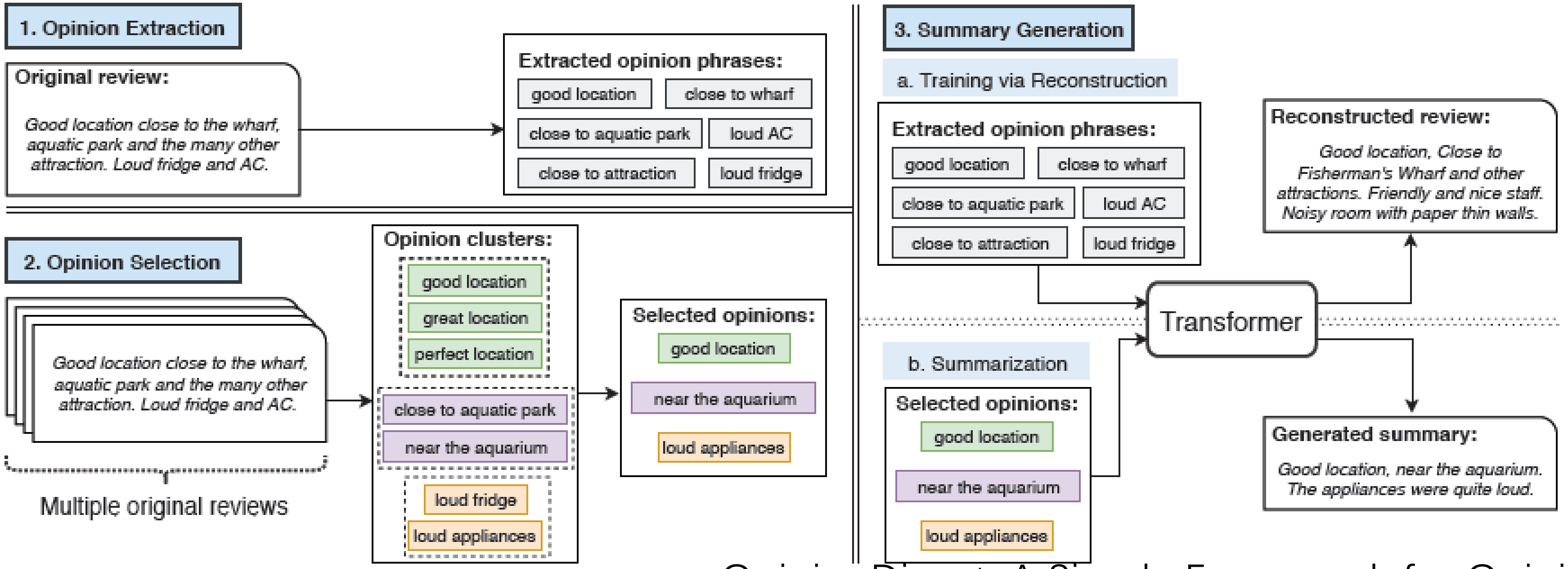
- Hierarchical query focused order-aware multi-document summarization model:
- Hierarchical Encoding
- Ordering Component
- Query Component



Data Augmentation for Abstractive Query-Focused Multi-Document Summarization

Opinion Summarization

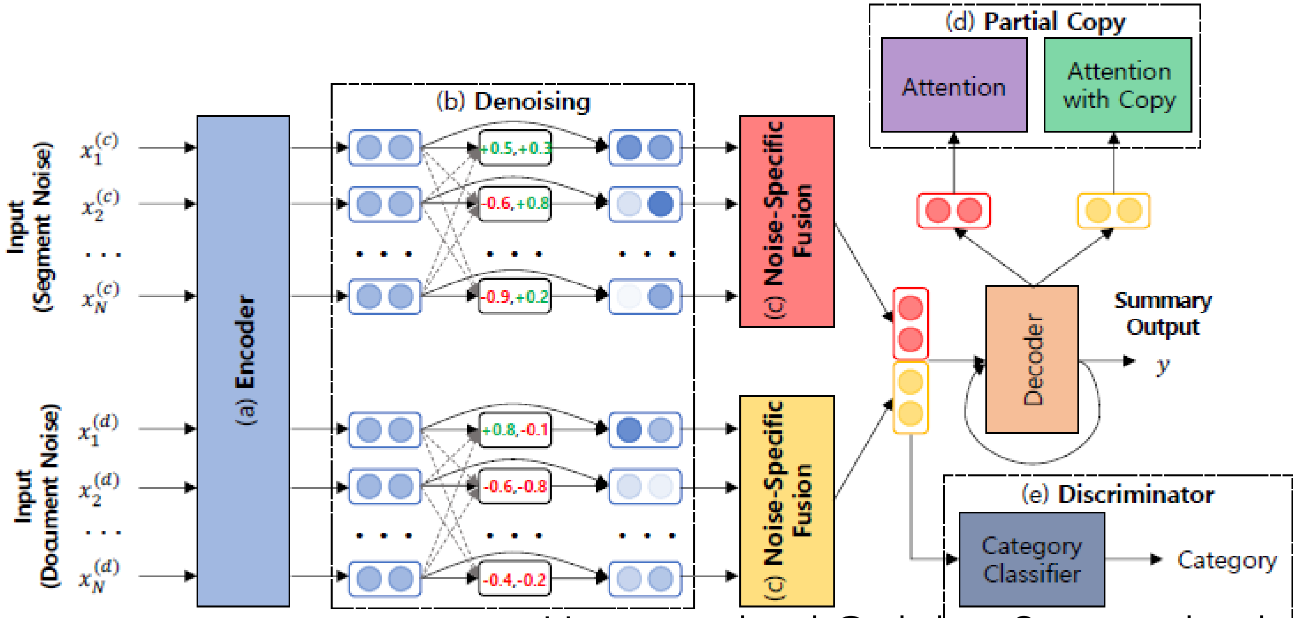
- Abstractive opinion summarization framework, which does not rely on gold-standard summaries for training
- Uses an Aspect-based Sentiment Analysis model to extract opinion phrases from reviews, and trains a Transformer model to reconstruct the original reviews from these extractions



OpinionDigest: A Simple Framework for Opinion Summarization

Opinion Summarization

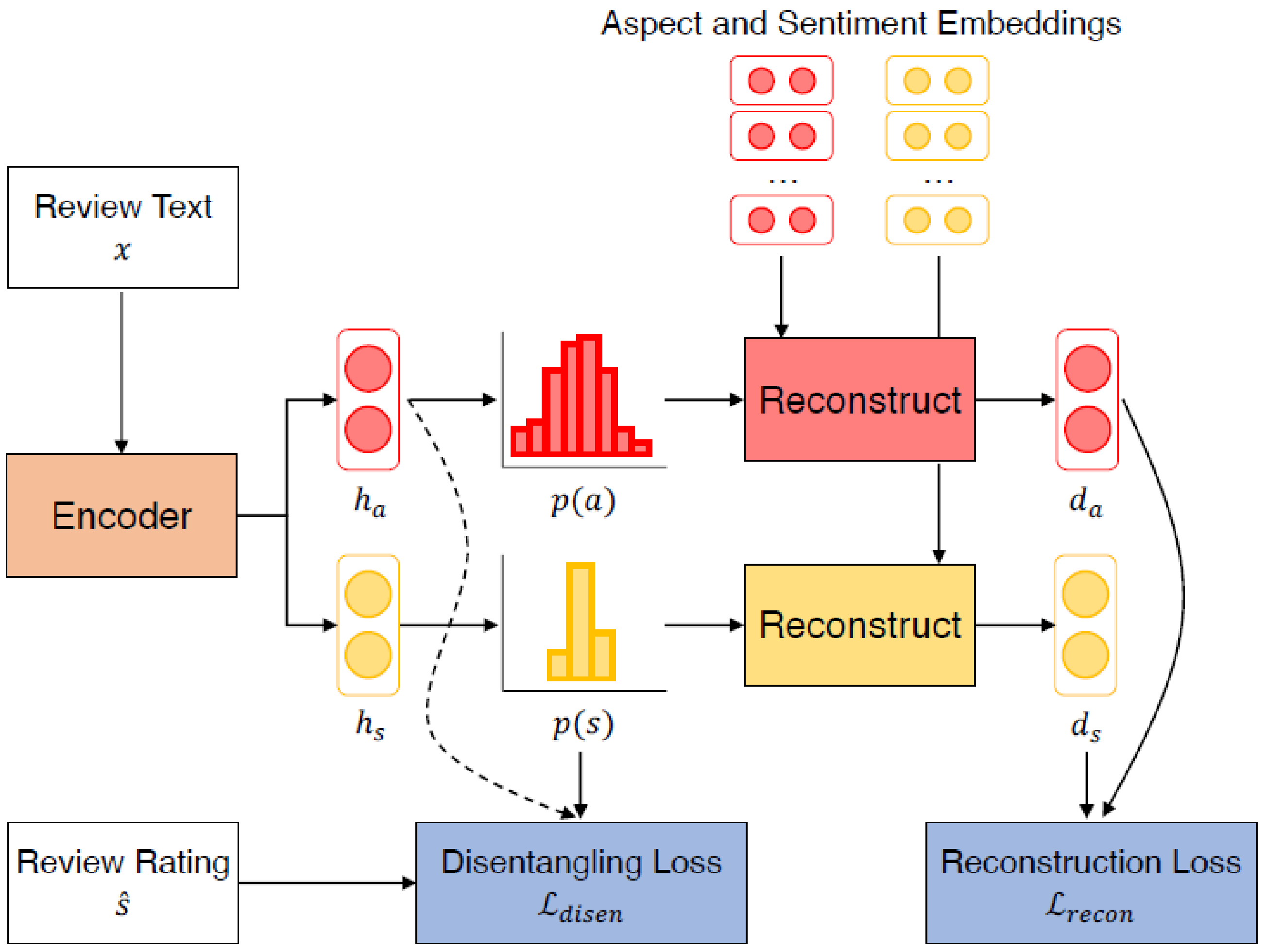
- Training data is not available and cannot be easily sourced
- Create a synthetic dataset from a corpus of user reviews by sampling a review, pretending it is a summary
- Generating noisy versions thereof which we treat as pseudo-review input



Unsupervised Opinion Summarization with Noising and Denoising

Opinion Summarization

- Training data is neither available nor can be easily sourced
- Explicitly incorporating *content planning* in a summarization model allows the creation of synthetic datasets



Unsupervised Opinion Summarization with Content Planning

Recent trends

- Multi-modal summarization
- Long document summarization
- Dialog summarization

Thanks!

Email: shengao@pku.edu.cn