A Hierarchical Encoding-Decoding Scheme for Abstractive Multi-document Summarization

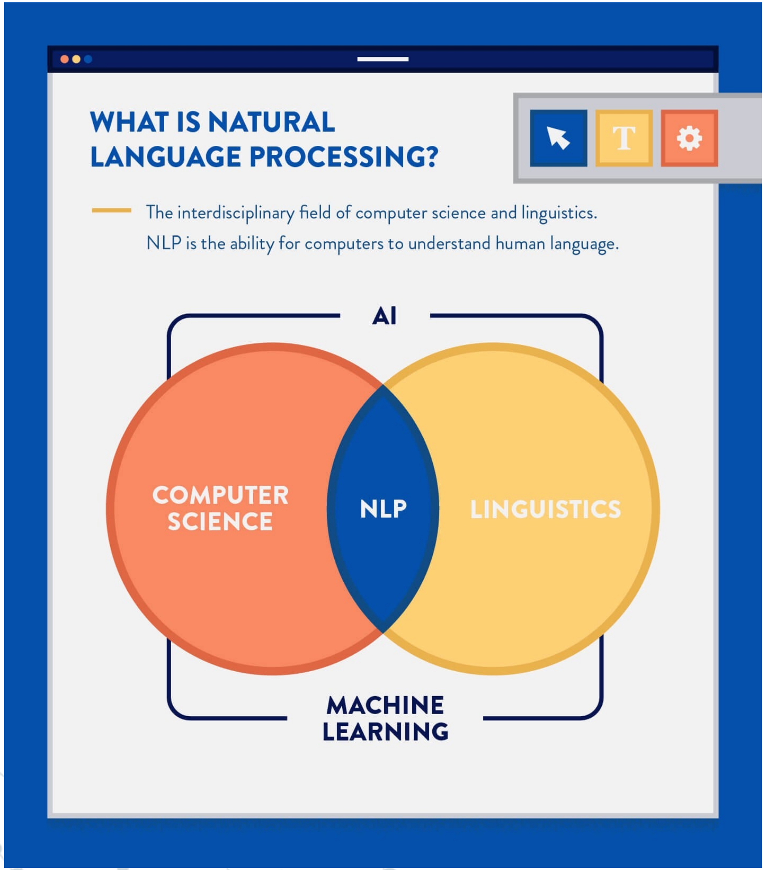
Presented by:
Kushan Hewapathirana – 229333P

Published in: EMNLP 2023: Main Volume: December 2023

APPLICATION DOMAIN:

Natural Language Processing – Multi-document Summarization

Introduction to Natural Language Processing



- Speech recognition
- Part of speech tagging
- Word sense disambiguation
- Named entity recognition
- Co-reference resolution
- Sentiment analysis
- Natural language generation

Introduction to Multi-document Summarization

MDS is a task that involves condensing information from multiple documents into a concise and coherent summary

Complex relationships between different documents, making it more intricate than single-document summarization (SDS)

Summarizing multiple documents with conflicting views is complex and almost impossible using extractive summarization

PROBLEM STATEMENT:

Inefficiency of previous methodologies in leveraging the capabilities of PLMs to enhance multi-document interactions

Challenges in Multi-document Summarization

Struggle to effectively leverage the benefits of pre-trained language models (PLMs) and may not adequately capture the nuanced relationships between multiple documents

Not generalize well across different domains, limiting their overall performance and ability to handle diverse sets of documents effectively

MDS poses challenges due to the intricate nature of handling cross-document information, which is more complex than in SDS

Bridging the Research Gap: Unique Contribution

Not leveraging the benefits of pre-trained language models (PLMs)

- By enforcing a hierarchical encoding-decoding scheme in both the encoder and decoder, the study aims to enhance the utilization of PLMs for MDS, which is a unique contribution in the field of text summarization

Inability to capture the nuanced relationships between multiple documents

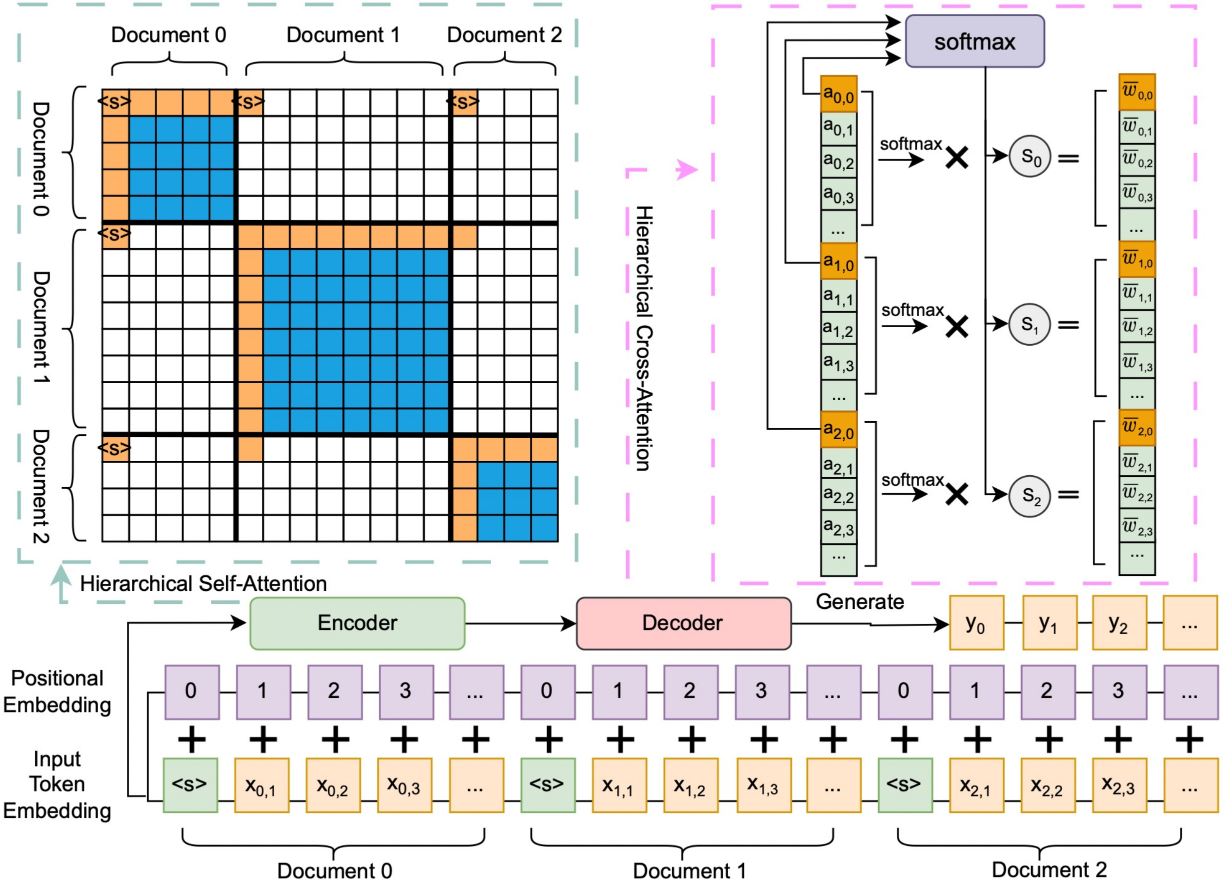
- The hierarchical approach in both the encoder and decoder proposed in the paper allows for a more comprehensive understanding and utilization of cross-document relationships inherent in MDS

Apply PLMs bluntly with concatenated source documents as a reformulated SDS task

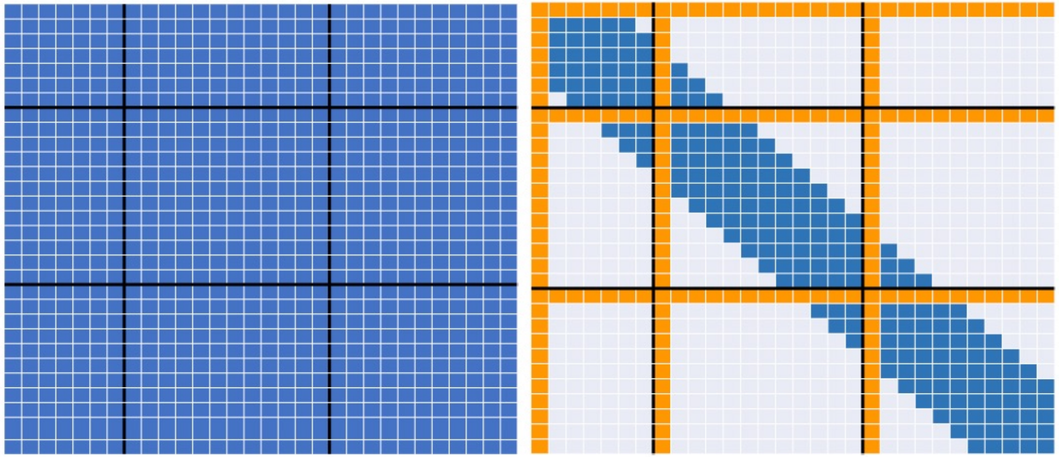
- Previous works either introduced specific MDS architectures or used PLMs directly for SDS tasks, without fully considering the complexities of MDS and leveraging the hierarchical structure for cross-document interactions that this study emphasizes

Methodology

Proposed Approach



Encoder Self-Attention Patterns in Different Attention Schemes



(a) Full attn (b) Global + local attn window

Dataset Statistics

Dataset	Instances	Docs	Len_{src}	Len_{tgt}	Train Steps
Multinews	56K	2.8	1793	217	130000
WCEP	10K	9.1	3866	28	15500
Multi-Xscience	40K	5.1	700	105	90000
Rotten Tomatoes	3K	100	2052	21	4500
MReD	6K	3.3	1478	120	10500
MReD+	6K	6.3	3069	120	10500
Film	37K	4.5	777	92	85000
MeanOfTransportation	10K	4.1	878	88	20000
Town	16K	4.7	582	52	37000
Software	15K	4.3	843	113	35000

EXPERIMENTS AND RESULTS

Experimental Setup

Data:

Datasets used: Multiple datasets i.e. Multinews, WCEP, Rotten Tomatoes

Baselines:

Fine-tuned Bart, LED, LongT5, PRIMERA, models.

Experimental Process:
Fine-tuned all evaluated models with cross-entropy loss on all datasets. Used Adam optimizer with a learning rate of 5e 5, and without any warm-up or weight decay.

Experimental Environment:

on single A100-80G GPU.

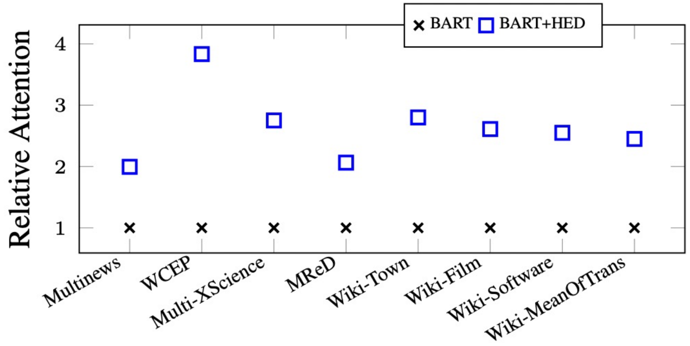
Test Results

		Multinews	WCEP	M-XSc	RT	MReD	MReD+	MeanOT	Town	Software	Film
System	Size	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L	R-1/R-L
LongT5	250M	46.4/24.5	43.4/35.3	27.0/15.0	26.0/20.5	32.0/20.1	**32.7**/20.6	**41.2/33.7**	60.2/56.7	**37.5**/28.4	42.4/35.5
BART(base)+HED	139M	**47.1/25.0**	**44.8/36.8**	**31.9/17.7**	**26.8/20.8**	**32.2/20.6**	**32.7/20.8**	40.6/**33.7**	**61.4/57.7**	37.2/**28.6**	**42.8/35.9**
LED	435M	50.1/25.0	46.5/37.6	31.2/16.6	27.3/20.7	33.0/19.1	34.3/20.3	45.4/35.1	62.3/**58.3**	42.1/28.8	44.8/35.7
PRIMERA*	447M	49.9/25.9	46.1/**37.9**	31.9/18.0	-	-	-	-	-	-	-
PRIMERA	447M	49.0/25.6	46.2/37.4	31.9/18.0	27.4/**21.1**	29.6/17.0	29.2/16.5	44.1/35.6	62.1/58.3	39.0/28.4	44.4/**36.9**
BART	406M	47.4/24.0	42.8/34.5	31.5/16.9	26.1/20.3	32.9/19.9	32.9/20.1	43.0/34.9	59.9/56.3	39.5/28.7	42.1/34.4
BART+HED	406M	50.0/25.8	46.4/37.8	32.1/17.6	27.3/**21.1**	33.9/**20.9**	34.0/**20.7**	43.5/35.2	61.9/57.7	40.5/**29.7**	43.8/36.3
BART-cnn+HED	406M	**51.1/25.9**	**47.0**/37.6	**34.7/18.6**	**27.6**/20.5	**34.1**/20.5	**34.5**/20.6	**46.1/35.4**	**62.8/58.3**	**42.9/29.7**	**45.9**/36.6

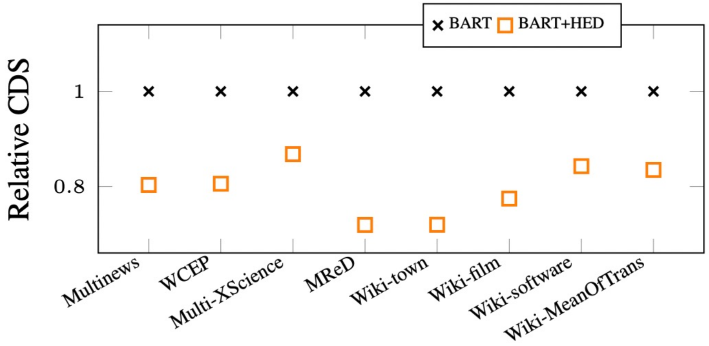
Test Results (Human Evaluations)

	Multinews					MReD				
model	Flu	Rel	Abs	Sal	Cov	Flu	Rel	Abs	Sal	Cov
BART	**0.510**	0.430	0.475	0.500	0.480	0.440	0.480	0.370	0.355	0.350
BART+HED	0.490	**0.570***	**0.525***	0.500	**0.520**	**0.550***	**0.520**	**0.630***	**0.645***	**0.650***

Document-level Attention Analysis



(a) Relative document self attention of "BART+HED" over "BART" in the **encoder**. For better visualization, we exclude the result for Rotten Tomatoes, which is 19.5.



(b) Relative cross-document standard deviation of "BART+HED" over "BART" in the **decoder**. We exclude the result for Rotten Tomatoes, which is statistically insignificant.

Content Analysis

System	Multinews	WCEP	M-XSc	RT	MReD	MReD+	MeanOT	Town	Software	Film
BART	0.71	4.47	**0.65***	1.19	0.82	0.84	0.24	**0.28**	0.24	0.36
BART+HED	**0.72**	**4.70***	0.46	**1.32**	**0.83**	**1.07***	**0.27***	0.27	**0.27***	**0.40***

Ablation Study

row	<s>	HAE	HAD	PR	Δ(R-1)	Δ(R-2)	Δ(R-L)
0	×	×	×	×	-	-	-
1	✓	×	×	×	+0.6	+0.7	+0.8
2	✓	✓	×	×	+0.9	+0.8	+0.8
3	✓	✓	×	✓	+1.0	+0.8	+0.7
4	✓	✓	✓	×	+0.9	+1.0	+0.9
5	✓	✓	✓	✓	+1.5	+1.3	+1.3

CONCLUSIONS

- The research paper focuses on abstractive MDS and introduces a hierarchical encoding-decoding scheme to enhance the performance of PLMs in this task.

- The proposed method enables effective fine-tuning of the PLM specifically for MDS without needing additional parameters, which simplifies the process.

- A key feature of the scheme is the use of global tokens for interactions between documents at both the encoding and decoding stages, allowing for a more comprehensive understanding of the multiple documents being summarized.

- By leveraging the generalizing capability of PLMs across various domains, the proposed approach can adapt well to different types of content and topics.

- Evaluation results from testing the approach on 10 different MDS datasets consistently show that it outperforms previous state-of-the-art models and even surpasses the performance of the PLM backbone itself.


THANK YOU…