Compressed Heterogeneous Graph for Abstractive Multi-document Summarization

Miao Li[#], Jianzhong Qi[+], Jey Han Lau[*]

School of Computing and Information Systems, The University of Melbourne

Background

- Abstractive multi-document summarization (MDS)
  - Input: topically related documents
  - Output: a concise and informative summary
- Summarizing multiple documents in an abstractive fashion

$$p(\hat{z}|\mathcal{D}) = \prod_{i=0}^{T} p(\hat{w}_i|\mathcal{D}, \hat{w}_0, \hat{w}_1, \ldots, \hat{w_{i-1}})$$

Related work and challenges

PLM-based MDS

- General-purpose PLMs, e.g.,
  - BART, Longformer, and T5
- Tailored-purpose PLMs, e.g.,
  - PEGASUS (Zhang et al. 2020a)
  - PRIMERA (Xiao et al. 2022)
- Drawbacks
  - Limited to learn **cross-document relationships** because of the flat concatenation of source documents

Graph-based MDS

- Only a handful models, e.g.,
  - Graphs of paragraphs (Li et al. 2020)
  - Hierarchical graphs based on the document structure (Jin et al. 2020)
- Drawbacks
  - Only leverage homogeneous graphs
  - without considering edge types of graphs
  - while the cluster of documents should be heterogeneous

Our solution of HGSum

The constructed heterogeneous graph

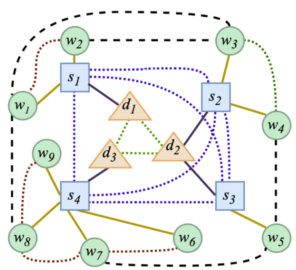

Figure 1: The structure of the heterogeneous graph given three documents in a document cluster: The orange triangles denote document nodes d, the blue quadrates denote sentence nodes s, the green circles denote word nodes w, and the line (or curve) segments between nodes denote edges. A detailed description of the graph is in the Preliminaries.

The architecture of HGSum

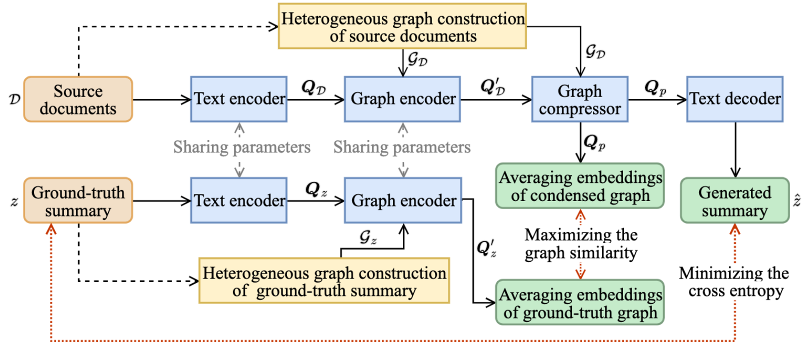

Figure 2: The HGSUM architecture: There are four main components: (1) text encoder (initialised using PRIMERA weights); (2) graph encoder; (3) graph compressor; and (4) text decoder (initialised using PRIMERA weights).

$$\mathcal{L} = \beta\mathcal{L}_{ce} + (1-\beta)\mathcal{L}_{gs} \qquad \mathcal{L}_{ce} = -\frac{1}{T}\sum_{i=1}^{T} w_i \log \hat{w}_i \qquad \mathcal{L}_{gs} = -\operatorname{sim}(\operatorname{avg}(\boldsymbol{Q}_p), \operatorname{avg}(\boldsymbol{Q}'_z))$$

Experiments

Main results

Model	Multi-News R-1	Multi-News R-2	Multi-News R-L	WCEP-100 R-1	WCEP-100 R-2	WCEP-100 R-L	Arxiv R-1	Arxiv R-2	Arxiv R-L
PEGASUS	47.70	18.36	43.62	42.43	17.33	32.35	44.21	16.95	38.87
LED	47.68	19.72	43.83	43.05	20.94	34.99	46.50	18.96	41.87
PRIMERA	<u>49.40</u>	<u>20.51</u>	<u>45.35</u>	<u>43.11</u>	**21.85**	<u>35.89</u>	<u>47.24</u>	<u>20.24</u>	<u>42.61</u>
MGSum	45.63	16.71	40.92	38.88	14.22	23.37	40.58	11.22	29.93
GraphSum	45.71	17.12	41.99	39.56	14.38	29.41	42.98	16.55	37.01
HGSum (our model)	**50.64**	**21.69**	**45.90**	**44.21**	<u>21.81</u>	**36.21**	**49.32**	**21.30**	**44.50**
Performance gain	+2.51%	+5.75%	+1.21%	+2.55%	-0.18%	+0.89%	+4.40%	+5.24%	+4.44%

Table 3: Model performance on summarizing Multi-News, WCEP-100, and Arxiv in terms of F1 of ROUGE scores. The best performance results are in boldface, while the second best is underlined.

✓ HGSUM outperforms most of the benchmark systems, demonstrating the effectiveness of incorporating a compressed heterogeneous graph for text summarization

Ablation study

Model	R-1	R-2	R-L	BScore
HGSum	50.64	21.69	45.90	87.38
w/o MGAT	48.87	20.32	43.21	87.08
w/o graph compressor	49.00	20.38	45.01	86.92
w/o multi-task training	48.10	20.30	44.24	86.85

Table 5: Results of ablation study on Multi-News.

Initialized by	R-1	R-2	R-L	BScore
random weights	18.99	27.86	16.88	79.32
LED	48.36	19.99	44.25	86.73
PRIMERA	50.64	21.69	45.90	87.38

Table 6: Summarization results of HGSum with different initialization on Multi-News.

[#] miao4@student.unimelb.edu.au [+] jianzhong.qi@student.unimelb.edu.au [*] laujh@unimelb.edu.au