VLSP 2022 | 9TH INTERNATIONAL WORKSHOP ON VIETNAMESE LANGUAGE AND SPEECH PROCESSING

Hanoi, November 26, 2022

Abstractive Multi-document Summarization

Task Overview

Mai-Vu Tran, Hoang-Quynh Le, Duy-Cat Can and Quoc-An Nguyen

ORGANIZERS

SPONSORS

Task description

VLSP2022 - Vietnamese abstractive multi-document summarization (Abmusu) task is to develop summarization systems that could create **abstractive summaries** automatically for **a set of documents** on a **topic**:

- Domain: Vietnamese news text.
- Input: Multiple news documents on the same topic.
- Output: Corresponding abstractive summary.

Data construction

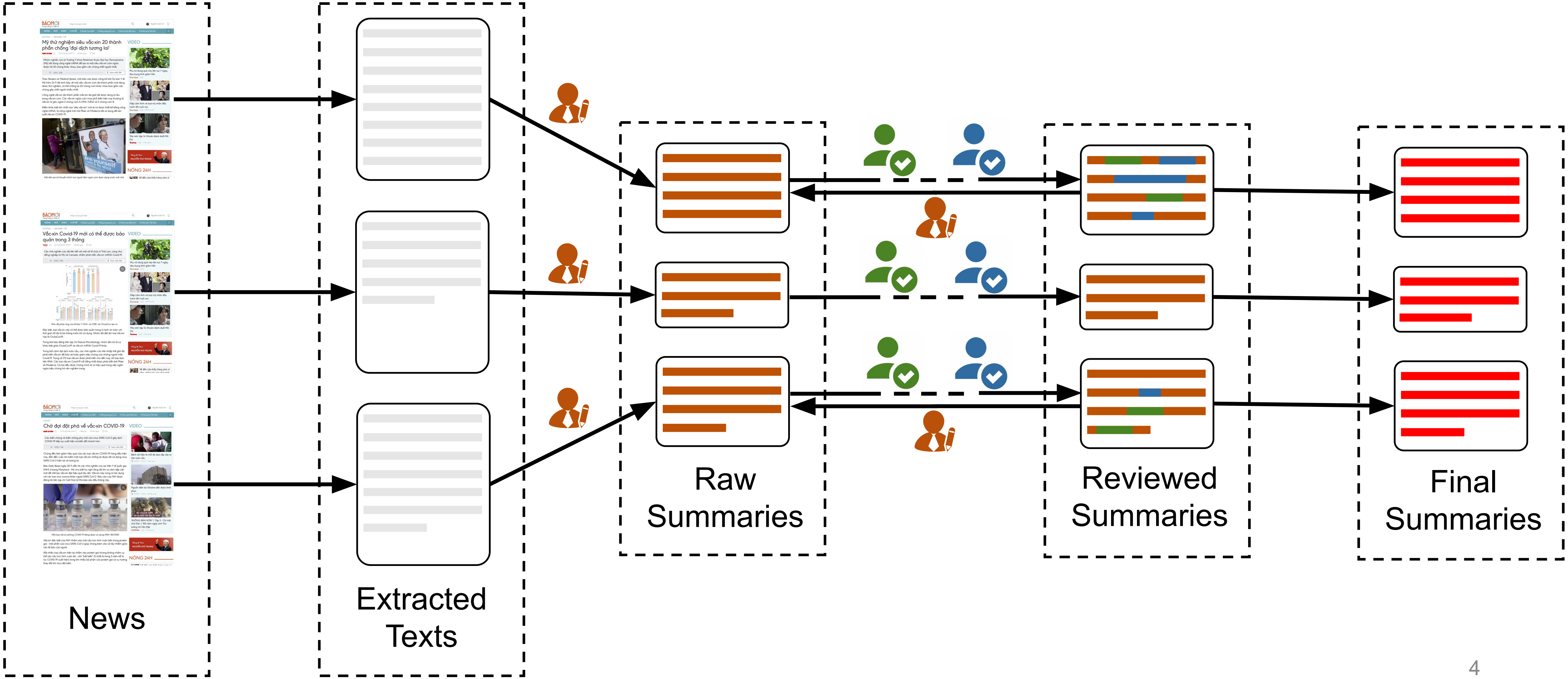
- Data construction has been done under the research project "Research and Development of Vietnamese Multi-document Summarization Based on Advanced Language Models" of Vietnam National University, Hanoi (QG.22.61)
- Sources: baomoi.vn
- 8 topics: Văn hóa - xã hội, pháp luật, kinh tế, khoa học – công nghệ, giải trí – thể thao, đời sống, thế giới, giáo dục
- Provided information:
  - Title, anchor text

    and body text of single documents
  - Summary
  - Category tag

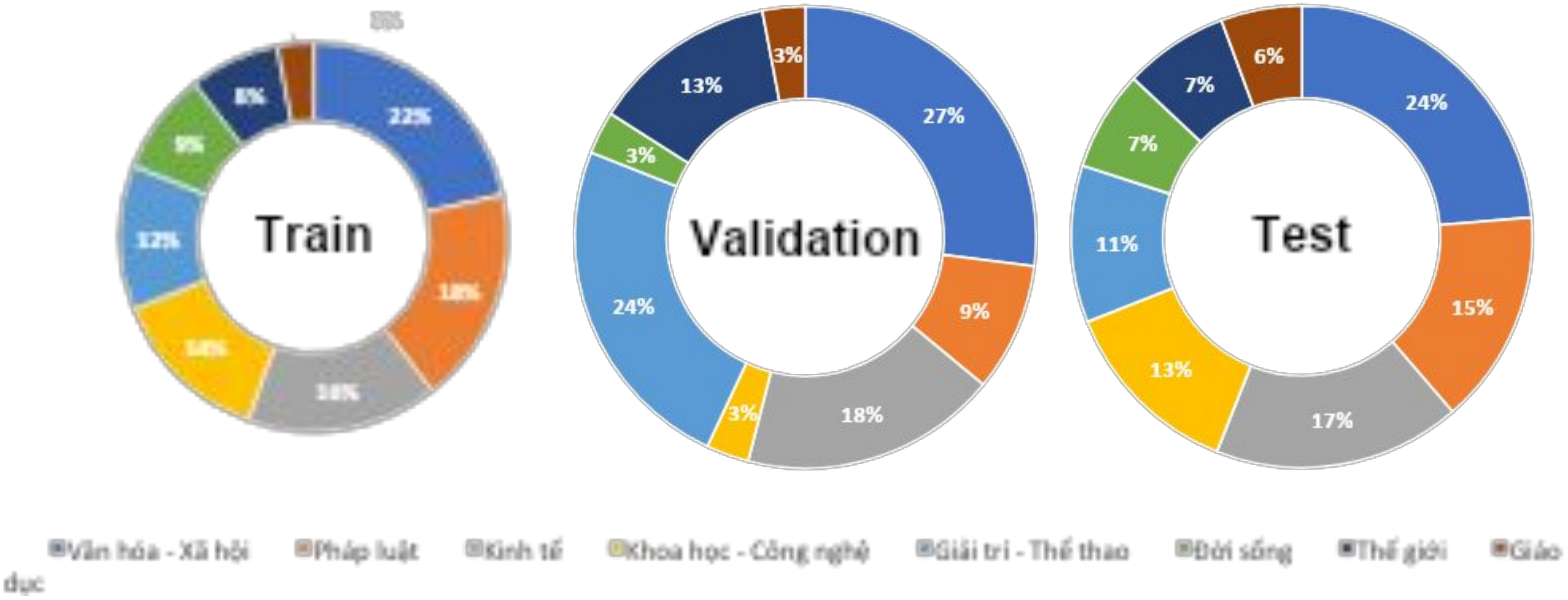
Number of	Training	Validation	Test
Clusters	200	100	300
Documents	621	304	914
Categories	8		

Numbers of clusters and documents by datasets

Data construction process



Data statistics



The statistics by categories

Data statistics

Aspects	Training	Validation	Test
Average			
Documents per Cluster	3.11	3.04	3.05
Tokens per Cluster	1924.75	1815.41	1762.40
Tokens per Raw text	619.88	597.17	578.46
Tokens per Anchor text	41.65	35.58	40.33
Tokens per Summary	168.48	167.68	153.05
Compression ratio			
Multi-document Summary	0.09	0.09	0.09

Average statistics and compression ratio

Participated teams

- Registered teams: 46 teams
- Participated teams: 28 teams
- Submitted results: 16 teams (287 submissions)
    - Domestic and international teams
    - Universities: VNU-HUS, VNU-UET, HUST, PTIT, etc.
    - Industries: Viettel, VinGroup, CMC, TopCV, VCCorp, etc.

Evaluations

- Platform: aihub.ml AI HUB
  - Public test (35 submission/7 days)
  - **Private test** (5 submission/3 days)
- Provided scores: ROUGE 1-P/R/F1, ROUGE 2-P/R/F1 and ROUGE L-P/R/F1
- Official evaluate score: **ROUGE 2-F1**

ROUGE-2 P = |Matched N-grams| / |Predict summary N-grams|

ROUGE-2 R = |Matched N-grams| / |Reference summary N-grams|

ROUGE-2 F = (2 x ROUGE-2 P x ROUGE-2 F) / (ROUGE-2 P + ROUGE-2 F)

Approaches and methods

4 baselines:

- Ad-hoc rule-based baseline:
    - First and last sentences
    - Rank 7 *(rank 10 public)*
- Anchor text-based baseline:
    - All anchor texts
    - Rank 19
- Extractive baseline:
    - Lexrank [1] + MMR [2]
    - Rank 6 *(rank 8 public)*
- Abstractive baseline:
    - ViT5 [3]
    - Rank 20

[1] Erkan, G., & Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, *22*, 457-479.
[2] Goldstein, J., & Carbonell, J. G. (1998, October). Summarization:(1) using MMR for diversity-based reranking and (2) evaluating summaries. In *Tipster text program phase iii: Proceedings of a Workshop held at Baltimore, Maryland, October 13-15, 1998* (pp. 181-195).
[3] Phan, L., Tran, H., Nguyen, H., & Trinh, T. H. (2022). ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation. *arXiv preprint arXiv:2205.06457*.

Proposed methods

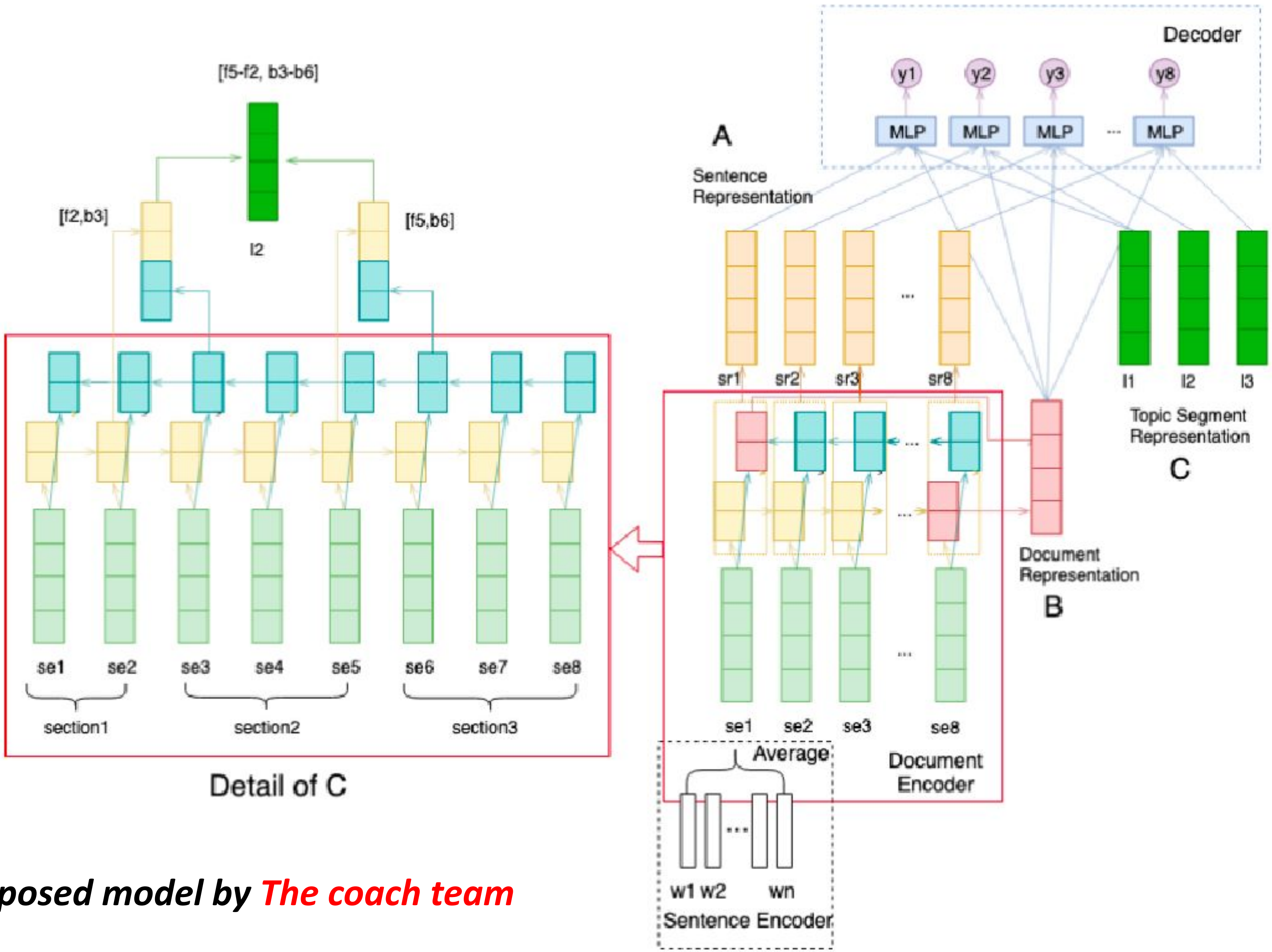
- Approaches:
  - Abstractive
  - Hybrid (Extracive x Abstractive)
- Extractive methods:
  - Similarity scoring: TF-IDF, Cosine, ...
  - Graph-based methods: Lexrank [1], Textrank [4], Pagerank [5]
  - Sentence classifier: LSTM, BERT [6]
  - Text correlation
  - …

[4] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411.
[5] Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.
[6] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT* (pp. 4171-4186).

Proposed methods

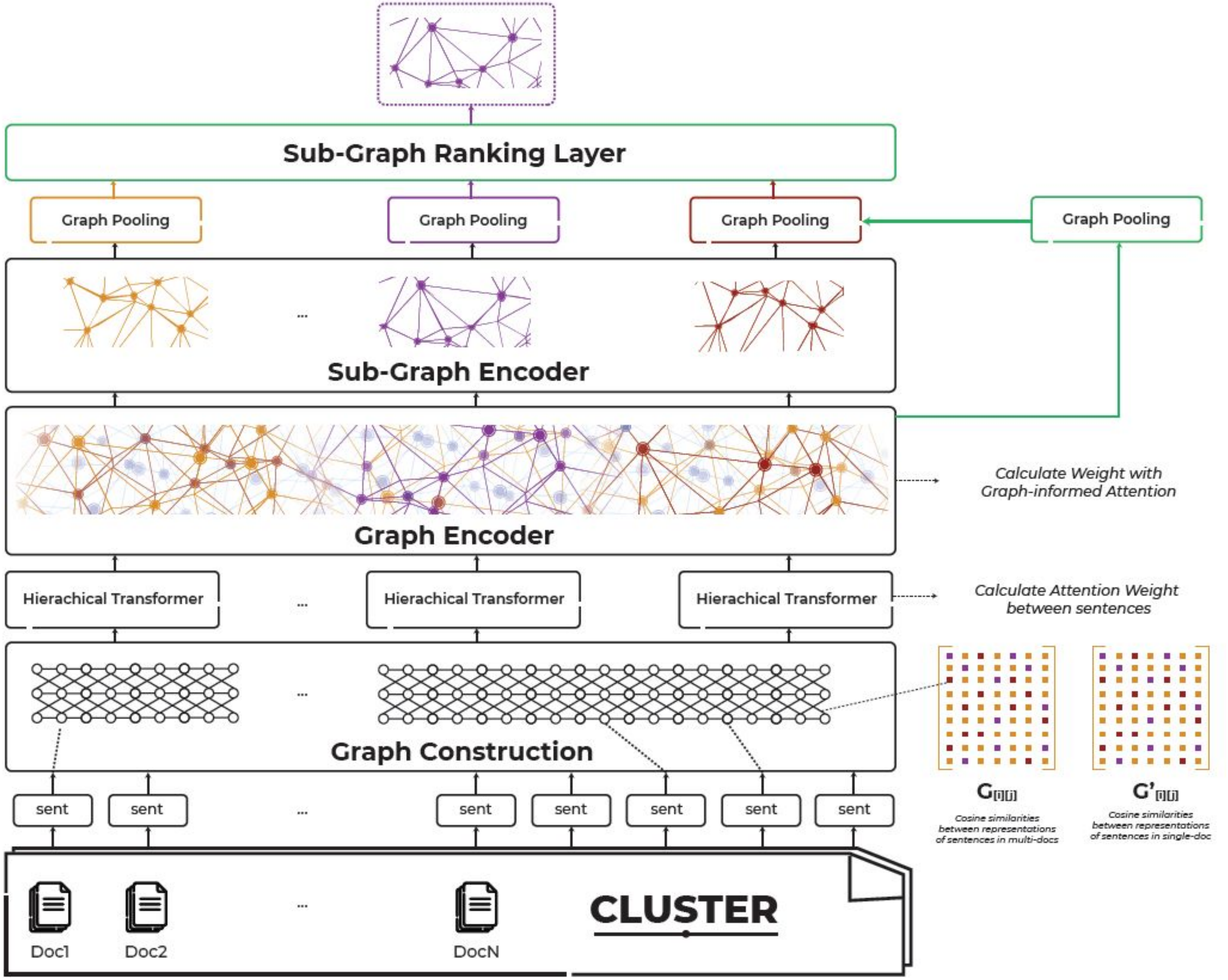
- Abstractive methods
  - Seq2seq model: ViT5 [3], BartPho [7], BRIO [8]
  - GL-LSTM [9]
  - Subgraph selection [10]
  - …
- Other techniques:
  - Maximal-Marginal Relevance [2]
  - Data augmentation
  - Post processing
  - …

[7] Nguyen Luong Tran, Duong Minh Le, and Dat Quoc Nguyen. 2021. Bartpho: Pre-trained sequence-tosequence models for vietnamese. arXiv preprint arXiv:2109.09701.
[8] Yixin Liu, Pengfei Liu, Dragomir Radev, and Graham Neubig. 2022. Brio: Bringing order to abstractive summarization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2890–2903.
[9] Xiao, W., & Carenini, G. (2019). Extractive summarization of long documents by combining global and local context. arXiv preprint arXiv:1909.08089.
[10] Moye Chen, Wei Li, Jiachen Liu, Xinyan Xiao, Hua Wu, and Haifeng Wang. 2021. Sgsum: Transforming multi-document summarization into sub-graph selection. arXiv preprint arXiv:2110.12645.



The proposed model by ***The coach team***

The proposed model by SGSUM team



Evaluation results - Public test

Rank	Team	ROUGE 2-F1	ROUGE 2-P	ROUGE 2-R	ROUGE 1-F1	ROUGE L-F1
1	**The coach**	**0.3150 (1)**	0.2492 (10)	0.4652 (2)	0.5168 (2)	0.4893 (1)
2	**LBMT**	**0.3149 (2)**	0.2566 (6)	0.4577 (3)	0.5178 (1)	0.4881 (2)
3	**The final year**	**0.2931 (3)**	0.2424 (12)	0.4137 (5)	0.5080 (3)	0.4733 (3)
4	Fcoin	**0.2875 (4)**	0.2978 (3)	0.2989 (10)	0.4902 (5)	0.4536 (5)
5	CIST AI	**0.2841 (5)**	0.2908 (4)	0.2949 (12)	0.4932 (4)	0.4547 (4)
7	vc-datamining	0.2680 (7)	0.1988 (18)	0.4696 (1)	0.4775 (8)	0.4507 (6)
8	*Extractive baseline*	*0.2650 (8)*	*0.2465 (11)*	*0.3268 (9)*	*0.4836 (7)*	*0.4421 (8)*
9	VNU Brothers	0.2586 (9)	0.3088 (1)	0.2365 (17)	0.4658 (11)	0.4227 (15)
10	*Rule baseline*	*0.2582 (10)*	*0.2521 (7)*	*0.2953 (11)*	*0.4640 (12)*	*0.4284 (13)*
19	*Anchor baseline*	*0.1931 (19)*	*0.2341 (13)*	*0.1753 (19)*	*0.4381 (18)*	*0.3928 (18)*
20	*Abstractive baseline*	*0.1457 (20)*	*0.3077 (2)*	*0.0995 (20)*	*0.3129 (20)*	*0.2797 (20)*

Evaluation results - Private test

Rank	Team	ROUGE 2-F1	ROUGE 2-P	ROUGE 2-R	ROUGE 1-F1	ROUGE L-F1
1	LBMT	**0.3035 (1)**	0.2298 (11)	0.4969 (1)	0.5067 (1)	0.4809 (1)
2	The coach	**0.2937 (2)**	0.2284 (12)	0.4463 (2)	0.4962 (2)	0.4701 (2)
3	CIST AI	**0.2805 (3)**	0.2629 (6)	0.3192 (6)	0.4876 (4)	0.4541 (4)
4	TheFinalYear	**0.2785 (4)**	0.2272 (13)	0.4040 (4)	0.4956 (3)	0.4612 (3)
5	NLP HUST	**0.2689 (5)**	0.2773 (4)	0.2829 (12)	0.4732 (6)	0.4373 (5)
6	*Extractive baseline*	**0.2625 (6)**	0.2464 (7)	0.3174 (8)	0.4772 (5)	0.4339 (6)
7	*Rule baseline*	**0.2611 (7)**	0.2634 (5)	0.2947 (10)	0.4627 (8)	0.4273 (8)
19	*Anchor baseline*	**0.1886 (18)**	0.2306 (10)	0.1734 (19)	0.4321 (17)	0.3869 (17)
20	*Abstractive baseline*	**0.1497 (19)**	0.3061 (1)	0.1025 (20)	0.3226 (20)	0.2895 (19)

Post-challenge panels

Post-challenge panels are opened on AIHUB for supporting your improvements:

- Unlimited number of submissions
- Register without approval

References

[1] Erkan, G., & Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, *22*, 457-479.

[2] Goldstein, J., & Carbonell, J. G. (1998, October). Summarization:(1) using MMR for diversity-based reranking and (2) evaluating summaries. In *Tipster text program phase iii: Proceedings of a Workshop held at Baltimore, Maryland, October 13-15, 1998* (pp. 181-195).

[3] Phan, L., Tran, H., Nguyen, H., & Trinh, T. H. (2022). ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation. *arXiv preprint arXiv:2205.06457*.

[4] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411.

[5] Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.

[6] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT* (pp. 4171-4186).

[7] Nguyen Luong Tran, Duong Minh Le, and Dat Quoc Nguyen. 2021. Bartpho: Pre-trained sequence-tosequence models for vietnamese. arXiv preprint arXiv:2109.09701.

[8] Yixin Liu, Pengfei Liu, Dragomir Radev, and Graham Neubig. 2022. Brio: Bringing order to abstractive summarization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2890–2903.

[9] Xiao, W., & Carenini, G. (2019). Extractive summarization of long documents by combining global and local context. arXiv preprint arXiv:1909.08089.

[10] Moye Chen, Wei Li, Jiachen Liu, Xinyan Xiao, Hua Wu, and Haifeng Wang. 2021. Sgsum: Transforming multi-document summarization into sub-graph selection. arXiv preprint arXiv:2110.12645.

…

Thank you for attending and listening!