



LTRC@MuP-2022: Multi-Perspective Scientific Document Summarization Using Pre-trained Generation Models

Ashok Urlana, Nirmal Surange, Manish Shrivastava
Language Technologies and Research Center, IIT Hyderabad, India
Scholarly Document Processing Workshop, COLING2022, Republic of Korea

Introduction

MuP2022 shared task

- Scientific documents with multiple summaries
- Leveraging multiple gold summaries to generate one multi-perspective summary.
- Explored several pretrained models such as BART, T5, ProphetNet.
- **Two Stage Fine-tuning approach:** First finetuning BART_{large} on SciTLDR dataset and then on the MuP dataset.
- We secured 3rd rank in the MuP-2022 shared task.

FineTuning PreTrained Models

Parameters	BART	T5	ProphetNet
Max source length	1024	1024	512
Max target length	150	128	128
Min target length	56	30	56
Batch Size	1	1	1
Epochs	2	10	1
Vocab Size	50265	32128	30522
Beam Size	4	4	5
Learning Rate	5e-5	1e-4	5e-5

Model	R-1	R-2	R-L	Avg R-f
Baseline	40.8	12.3	24.5	25.8
BART _{large} cnn	40.68	12.47	24.99	26.05
DistilBART cnn	39.36	11.79	24.47	25.21
BART _{base} cnn	39.12	11.42	23.8	24.78
T5 _{base}	38.35	11.26	24.64	24.75
ProphetNet	38.15	11.45	24.25	24.62
BART _{base}	38.53	11.39	23.92	24.61
ProphetNet cnn	37.59	10.91	24.09	24.2
DANCER + BART	33.07	9.06	18.2	20.11
BART + Two-stage	32.51	6.82	20.64	19.99

Experimenting with Input Data

To investigate the contribution of various sections of the scientific documents in the target summaries, we created the following four categories of source content:

1. **Introduction**
2. **Abstract + Introduction**
3. **Abstract + Introduction + Conclusion**
4. **Abstract + Conclusion**

	Train & Val Data				Test Data								
	R-1	R-2	R-L	Avg R-f	1	2	3	4	1	2	3	4	
1. Introduction	40.68	12.47	24.99	26.05	✓					✓			
2. Abstract + Introduction	40.67	12.5	24.93	26.03	✓						✓		
3. Abstract + Introduction + Conclusion	40.47	12.29	24.76	25.84			✓			✓		✓	
4. Abstract + Conclusion	40.34	12.28	24.79	25.8				✓					✓
	40.33	12.28	24.75	25.79		✓				✓			
	40.39	12.25	24.73	25.79			✓				✓		
	40.23	12.32	24.77	25.77	✓								✓
	40.23	12.17	24.6	25.67				✓		✓			
	40.1	12.25	24.63	25.66		✓			✓				
	40.22	12.13	24.54	25.63	✓				✓				

We fine-tuned and tested BART_{large} cnn with all combinations of the above categories of data (consuming one as train and the other as test set). We found that, among these combinations, the model trained with **Introduction** and tested with **Abstract + Introduction** performed the best.

Analysis & Conclusion

- Among all, the BART_{large} cnn model performed the best.
- While this task considers summaries from multiple reviewers as different “perspectives”, most of these summaries cover only the major contributions of the paper.
- These summaries, though diverse in their construction, do not look at the research paper from different points-of-view.
- We see a validation of this claim in our experiments where a model trained on “introduction” section alone outperforms all other combinations.