

# GraM-SUM: A Graph-Based Model for Book-Length Summarization

Romain L. Hardy  
W266 Section 001  
August 1, 2022

# Overview

- Summarization is well studied in NLP
- Most work focuses on short texts: newspaper articles, academic papers, short stories, etc.
- Book-length summarization is difficult
  - Model must capture long-term relationships across thousands of tokens
  - Transformers cannot process entire books in memory
- Solution: Break things into smaller pieces!



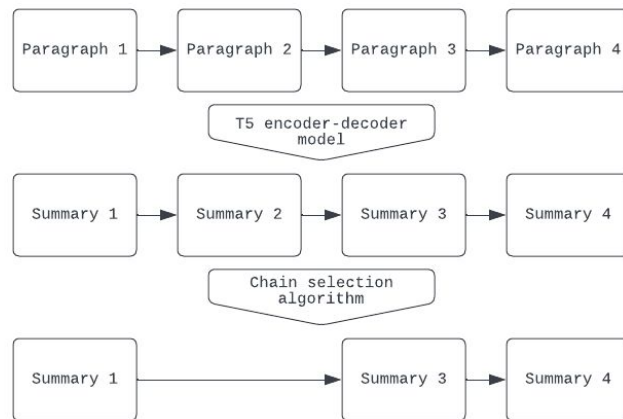
# Data and Metrics

- Main dataset: BookSum (2021)
  - Publicly available novels, plays, and stories from *Project Gutenberg*
  - Reference summaries scraped from independent sources like *SparkNotes* and *Shmoop*
  - 42,753 paragraph-level, 12,293 chapter-level, and 436 book-level text-summary pairs
- Evaluation metric: ROUGE
  - Use word-level stemming when computing scores



# Approach

- Fine-tune an encoder-decoder model on paragraph-level summaries from BookSum
- To create chapter/book-level, combine paragraph/chapter-level summaries
- Baseline: Rank paragraph summaries by perplexity, choose top  $k$  summaries
- GraM-SUM: Choose the chain of  $k$  summaries that is optimal in terms of progression, diversity, and importance

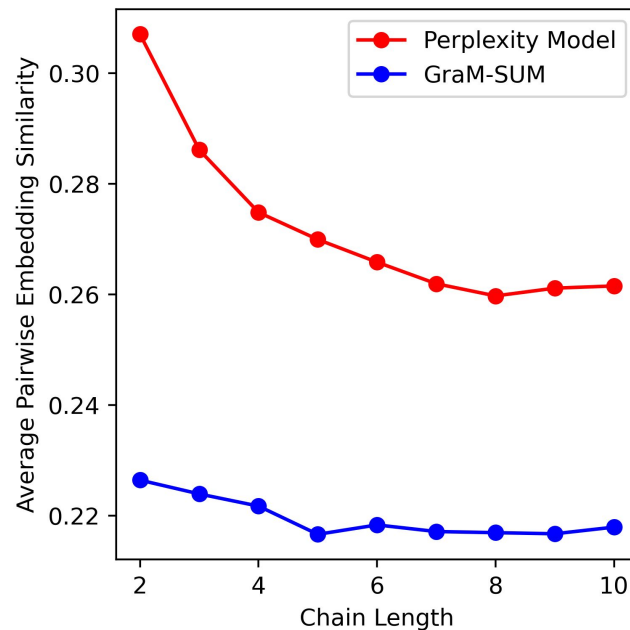


# Results

$k$	Perplexity (Chapter)			GraM-SUM (Chapter)		
	R-1 <sub>F1</sub>	R-2 <sub>F1</sub>	R-L <sub>F1</sub>	R-1 <sub>F1</sub>	R-2 <sub>F1</sub>	R-L <sub>F1</sub>
2	20.79	3.85	12.15	22.67	4.13	12.81
3	26.34	4.98	14.24	27.84	5.21	14.73
4	30.14	5.80	15.60	31.38	6.04	15.97
5	32.68	6.41	16.43	33.69	6.59	16.69
6	34.48	6.85	16.94	35.34	7.01	17.15
7	35.89	7.21	17.35	36.58	7.34	17.51
8	36.89	7.46	17.63	37.43	7.58	17.75
9	37.71	7.72	17.89	38.14	7.81	17.98
10	38.33	7.91	18.06	38.53	7.92	18.05

$k$	Perplexity (Book)			GraM-SUM (Book)		
	R-1 <sub>F1</sub>	R-2 <sub>F1</sub>	R-L <sub>F1</sub>	R-1 <sub>F1</sub>	R-2 <sub>F1</sub>	R-L <sub>F1</sub>
2	23.50	4.73	11.47	29.33	5.51	12.80
3	29.84	6.09	13.46	35.74	6.87	14.39
4	34.63	7.09	14.57	38.73	7.70	15.00
5	37.37	7.60	15.02	40.19	8.24	15.19
6	39.37	8.02	15.50	41.24	8.58	15.37
7	40.32	8.29	15.55	41.37	8.74	15.27
8	40.79	8.59	15.55	41.31	8.80	15.10
9	40.82	8.73	15.59	40.78	8.85	14.97
10	40.50	8.40	15.70	40.57	8.80	15.45



# Conclusions

- GraM-SUM generally outperforms the perplexity baseline at chapter and book levels
- Sentence-level embeddings are more diverse
- GraM-SUM is much slower than the perplexity model
- High-level summaries often seem disjointed



# Future Work

- Use a more powerful transformer model as the base for GraM-SUM
- Other graph-based features
- Add explicit penalty for choosing long summaries when constructing chains
- Evaluate using other metrics besides ROUGE



# Appendix: Example Chapter-Level Summary

Lodovico Moro is a son of Francesco Sforza, who married Beatrice d'Este, who ruled over Milan from 1494 to 1500, and died in 1510. Louis XII, King of France, born 1462, died 1515, and Charles VIII, king of France born 1470, died 1498. Louis was brought into Italy by the ambition of the Venetians, who wanted to obtain half the state of Lombardy by his intervention. Louis XII divorced his wife, Jeanne, daughter of Louis, and married Anne of Brittany, widow of Charles VIII, in order to retain the Duchy for the crown. He was Georges d'Amboise, created a cardinal by Alexander VI. Born 1460, died 1510. All states, all powers, that have held and hold rule over men have been and are either republics or principalities. The new are either entirely new, as was Milan to Francesco Sforza, or they are, as it were, members annexed to the hereditary state of the prince who has acquired them. Such dominions thus acquired are either accustomed to live under a prince, or to live in freedom; and are acquired either by the arms of the Prince himself or of others, or else by fortune or by ability. The Romans, in the countries that they annexed, observe closely these measures; they sent colonies and maintained friendly relations with the minor powers, without increasing their strength; they kept down the greater, and did not allow strong foreign powers to gain authority. They were kept friendly by them, the kingdom of Macedonia was humbled, Antiochus was driven out; yet the merits of the Achaeans and Aetolians never secured for them permission to increase their power.

