

What's in a Summary? Laying the Groundwork for Advances in Hospital

Course Summarization

Griffin Adams, Emily Alsentzer, Mert Ketenci, Jason Zucker, Noémie Elhadad Elhadad



Hospital Course Summarization

- Given the clinical documentation available for a patient hospitalization, synthesize the hospital course in a faithful and concise fashion.

1. What was done to the patient?
2. Why it was done?
3. What still needs to be done?

Dataset

- Materials come from all hospitalizations between 2010 and 2014 at Columbia University Irving Medical Center. Brief Hospital Course section of discharge summary serves as a proxy reference.

	Variable	Value	STD
Global	# Patients	68,936	
	# Admissions	109,726	N/A
	# Source Notes	2,054,828	
Per Adm.	Length of Stay	5.8 days	9.0
	# Source Notes	18.7	30.1
	# Source Sentences	1,061.2	1,853.6
	# Source Tokens	11,838.7	21,506.5
	# Summary Sentences	17.8	16.9
	# Summary Tokens	261.9	233.8
Per Sent.	# Source Tokens	10.9	12.4
	# Summary Tokens	14.5	11.5
Ratio	Word Compression	42.5	164.6

Table 1: Basic Statistics for CLINSUM. Value is the total for Global, and average for 'Per Admission' and 'Per Sentence'. STD is standard deviation.

Extractive Baselines

Extractive Baseline	ROUGE-1			ROUGE-2		
	Recall	Precision	F1	Recall	Precision	F1
RANDOM	0.16	0.24	0.17	0.04	0.03	0.03
LEXRANK	0.18	0.21	0.18	0.05	0.05	0.05
CLINNEUSUM	0.36	0.25	0.27	0.14	0.1	0.11
ORACLE TOP-K	0.28	0.52	0.32	0.16	0.32	0.19
ORACLE GAIN	0.43	0.63	0.5	0.26	0.42	0.3
ORACLE SENT-ALIGN (SA)	0.48	0.61	0.52	0.3	0.33	0.31
ORACLE RETRIEVAL	0.51	0.70	0.58	0.25	0.28	0.29
ORACLE SA + RETRIEVAL	0.6	0.76	0.66	0.4	0.49	0.43

Table 2: Performance of different sentence selection strategies on CLINSUM.

ClinNeusum - a variant of the Neusum model

Oracle Top-K - highest scoring ROUGE sentences

Oracle Gain - Greedily take source sentence with highest relative ROUGE

Oracle Sent-Align - per sentence Oracle Top-K

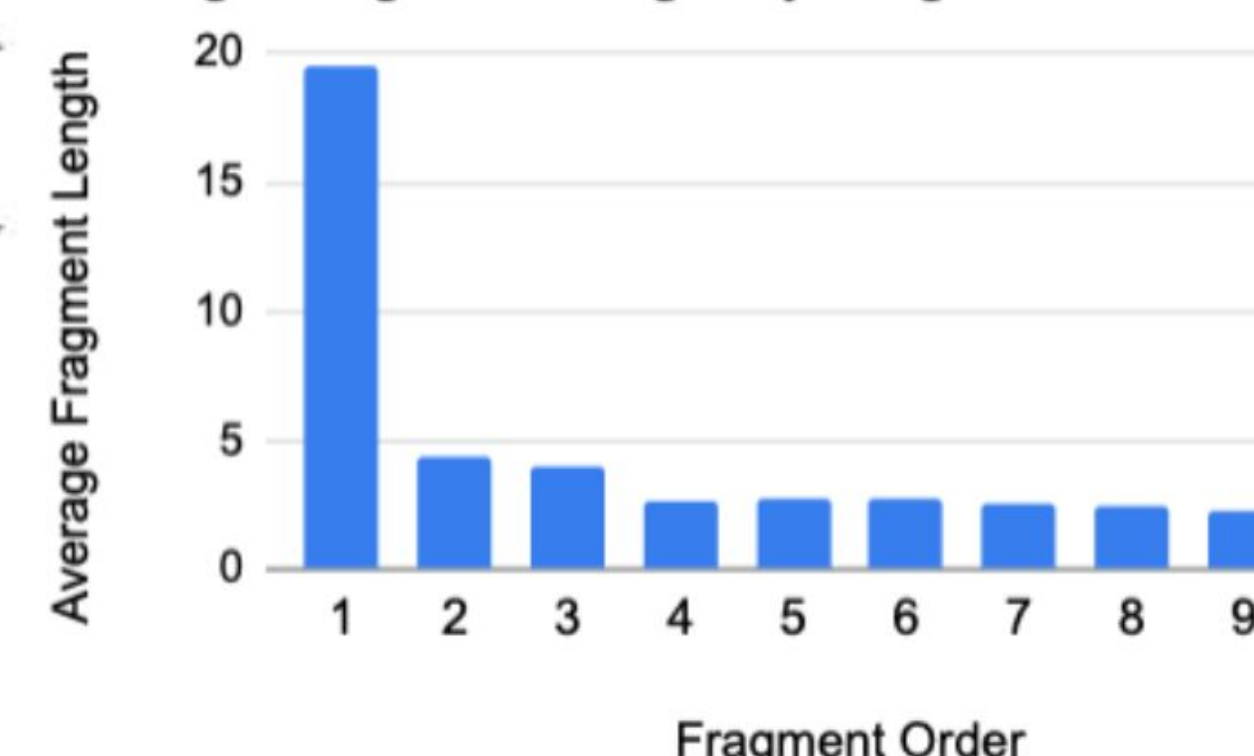
Oracle Retrieval - find best match from other summary for each sentence

Oracle SA + Retrieval - Oracle combination

Summaries abstractive despite copy-paste

- First sentence is often Copy-Paste yet rest is abstractive. An extractive model gets worse at selecting sentences as it becomes abstractive.

Average Fragment Length by Fragment Order



Extractive Step	Ground Truth Rank	
	Average	Median
1	28	7
2	69	22
3	74	31
4	79	39
5	76	42
> 5	80	60

Table 4: Rank of selected sentence vis-a-vis oracle rank at each extraction step. A perfectly trained system would have a ground-truth of 1 at each step.

Summaries concise yet comprehensive

- Packed with medical entities, which are well-distributed across the source notes. Relevant relations are often not explicit.

Organization & Style Transfer

- The Brief Hospital Course section is more problem-oriented than preceding text.

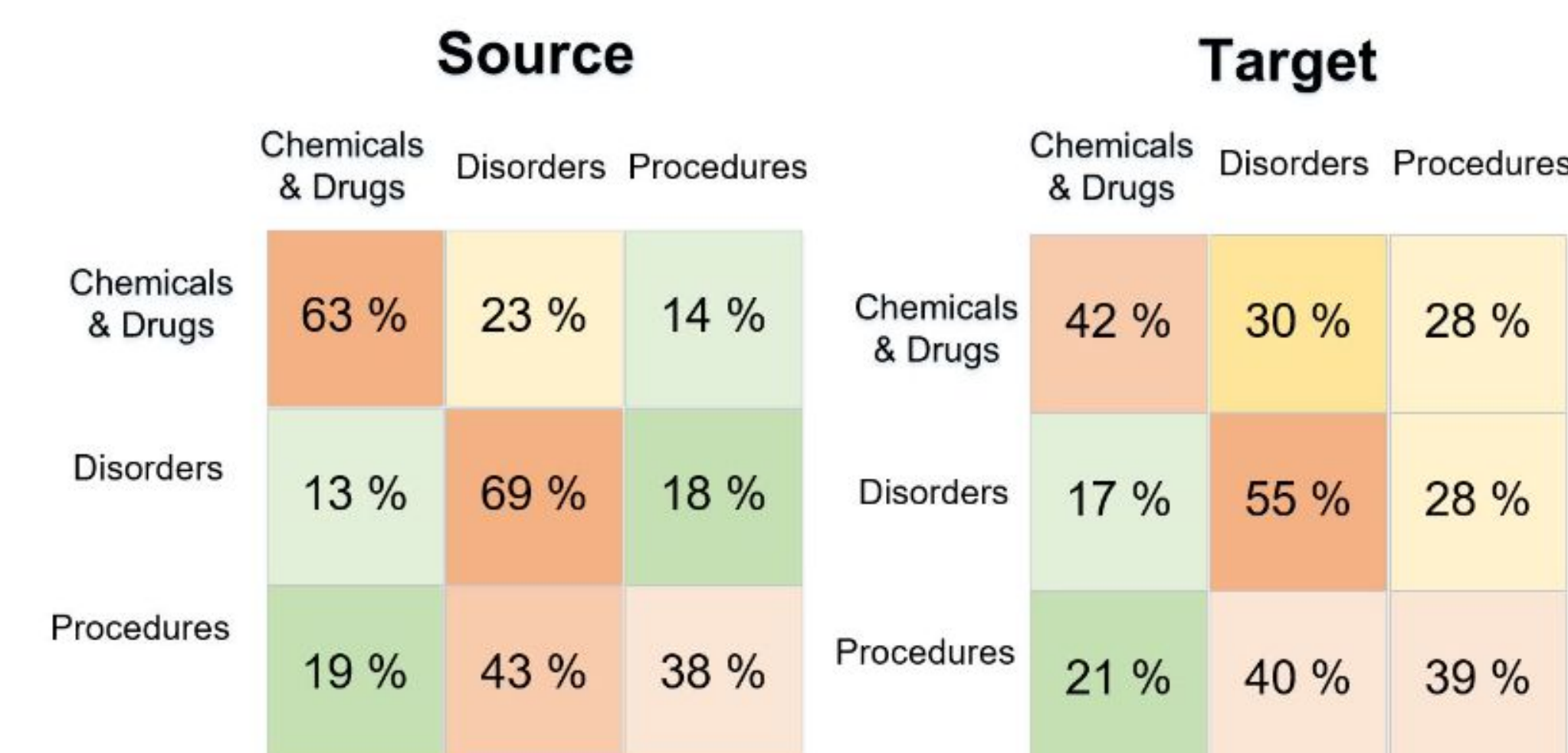


Figure 7: Entity Transition Matrices for source notes and target summaries. Summaries have fewer clusters of semantically similar entities, indicating that entity mentions are woven into a problem-oriented summary.

Low Lexical Cohesion

- Summaries exhibit frequent, abrupt topic shifts.

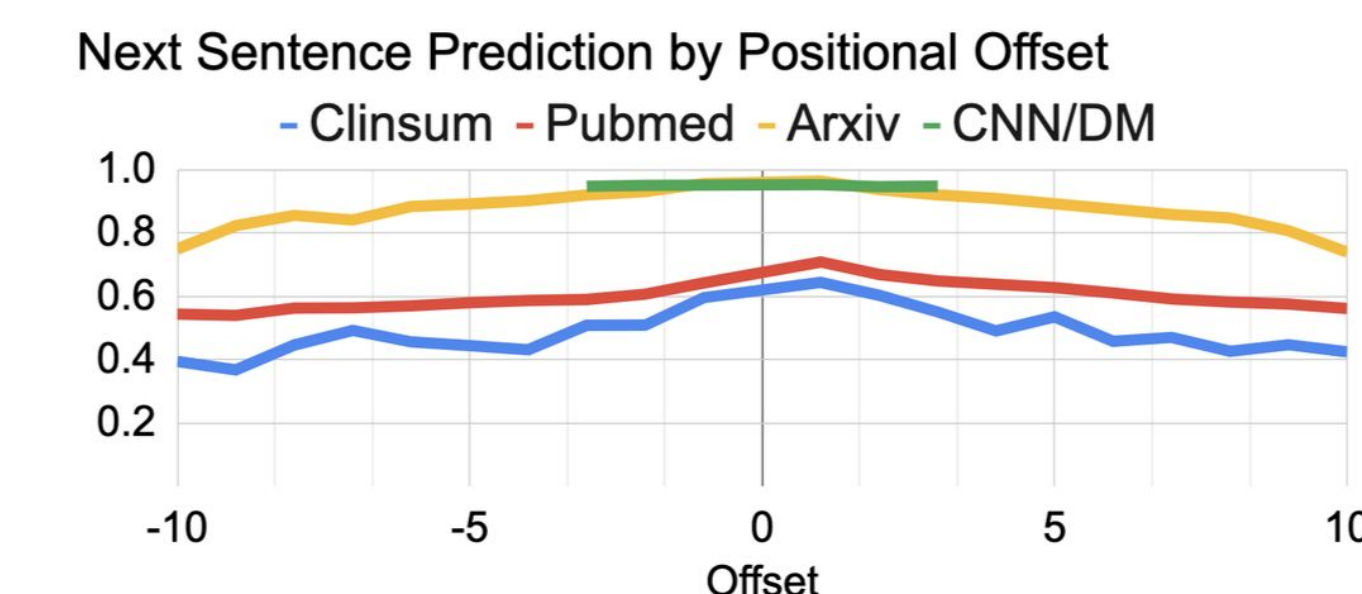


Figure 8: NSP logit by relative position of the next sentence across summaries for several datasets. An offset of 1 corresponds to the true next sentence.