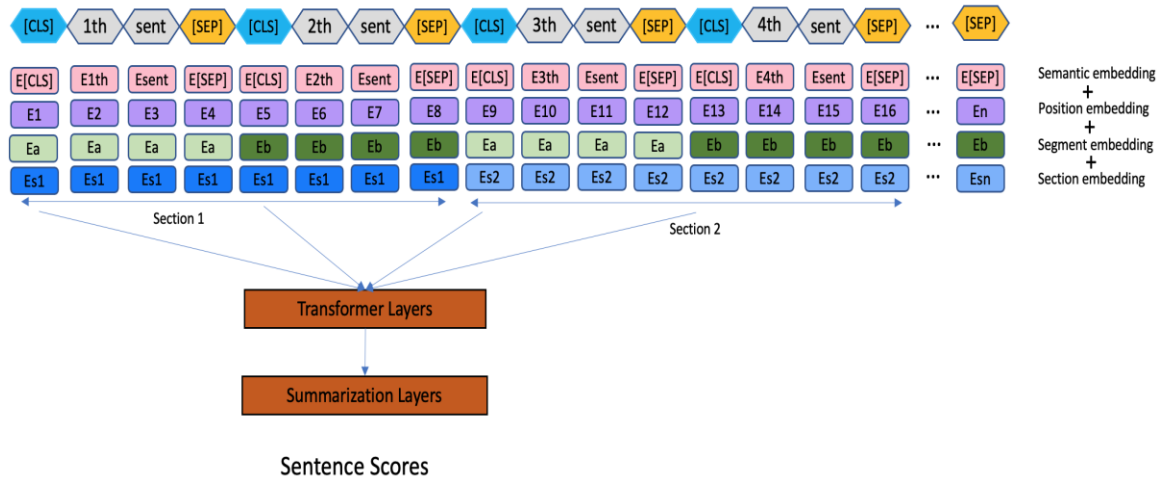


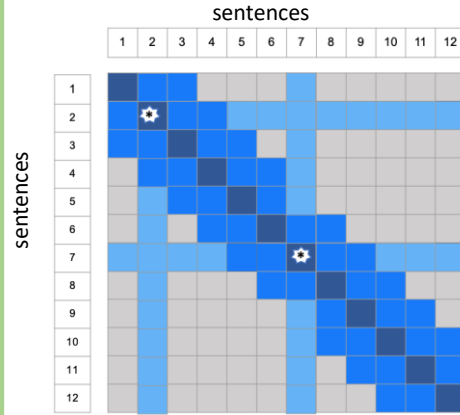
Language Model

- This language model is designed for summarization of long documents with upto 500 sentences
- Presentation slides are used as gold standard summaries
- Sparse attention models are used to represent inter sentence relations
- Section embedding differentiate sections of the document such as introduction/method/results



Attention Mechanism

- We apply a sparse inter sentence attention mechanism to identify important sentences from document
- Each sentence attends to a number of sentence before and after itself
- Some random sentences attend globally to all other sentences



Local Window Size	Global Ratio(%)	ROUGE-1	ROUGE-2	ROUGE-L
6	-	56.854	19.692	41.210
10	-	58.854	20.392	41.810
20	-	59.06	20.77	42.00
30	-	58.989	20.664	42.031
40	-	58.97	20.44	41.91
50	-	59.408	21.099	42.232
40	20	59.47	21.11	42.34
40	40	59.72	21.45	42.77
50	20	59.829	21.479	42.973
50	40	59.714	21.498	43.057

Summarization features

$$E_{length} = ReLU(Linear(Embedding(length[i])))$$

$$E_{position} = ReLU(Linear(Embedding(i)))$$

$$E_{section} = ReLU(Linear(Embedding(section[i])))$$

$$Correlation = \tanh(E_{sents} \times W_c \times E_{sents}^T)$$

$$E_{correlation} = ReLU(Linear(Correlation \times E_{sents}))$$

$$p(y_i) = Linear(E_{sent} + E_{length} + E_{position} + E_{section} + E_{correlation} + E_{saliency})$$

Results

comparing our model to baseline models

Table 3: Comparison with baselines based on ROUGE recall.

MODEL	ROUGE-1	ROUGE-2	ROUGE-L
Lead20%	37.68	6.62	15.90
TextRank [13]	38.87	9.28	19.75
SummaRuNNer [15]	45.04	11.67	23.03
BART (section-based)	46.34	11.14	29.85
T5 (section-based)	44.72	10.23	29.63
BERTSUM	52.34	15.06	36.87
SciBERTSUM	59.714	21.498	43.057