Incorporating Distributions of Discourse Structure for Long Document Abstractive Summarization

Dongqi Pu Yifan Wang Vera Demberg

Department of Computer Science
Department of Language Science and Technology
Saarland Informatics Campus, Saarland University, Germany

UNIVERSITÄT DES SAARLANDES

European Research Council
Established by the European Commission

Abstract

We introduce the 'RSTformer', a novel summarization model that incorporates both the **types** and **uncertainty** of rhetorical relations. Our RST-attention mechanism, based on document-level rhetorical structure, extends the Longformer model. Through evaluation, our proposed model outperforms existing models, as demonstrated by its notable performance on multiple metrics and human evaluation.

Introduction

For writing a concise and coherent summary of a long document, it is crucial to identify the salient information and comprehend the intricate connections between its different components. Rhetorical Structure Theory (RST) serves as a discourse framework that is designed to articulate the interrelationships among sentences at the document level. Within RST, two types of relations are distinguished: **paratactic relations**, where both segments hold equal importance, and **hypotactic relations**, which establish a hierarchical structure with a central 'nucleus' and a less central 'satellite' within the discourse. RST has proven effective in summarization tasks. However, there are two main problems with the current literature:

- RST relation types are overlooked
- Solely relying on the 1-best RST results

Method

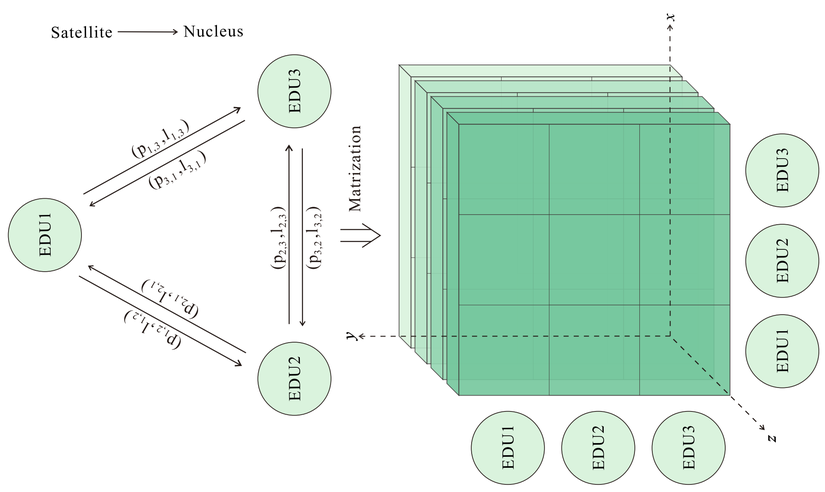

Figure 1. Labeled discourse distributions

Our methodology enhances discourse-injected text summarization by retaining parser uncertainty and exploiting labeled probability distributions. We model the discourse-driven seq2seq summarization task as:

$$P(t|s,d) \approx \prod_{i=1}^{T} P(t_i|t_{<i}, \text{encode}(s,d)) \quad (1)$$

where s, t, and d are the source, target sequence, and discourse representation, respectively. We then develop a tensor representation for the discourse structure, transforming all potential RST relations into a three-dimensional Labeled Discourse Distribution (LDD) tensor, as depicted in Figure 1, yielding a compact representation for the Longformer model.

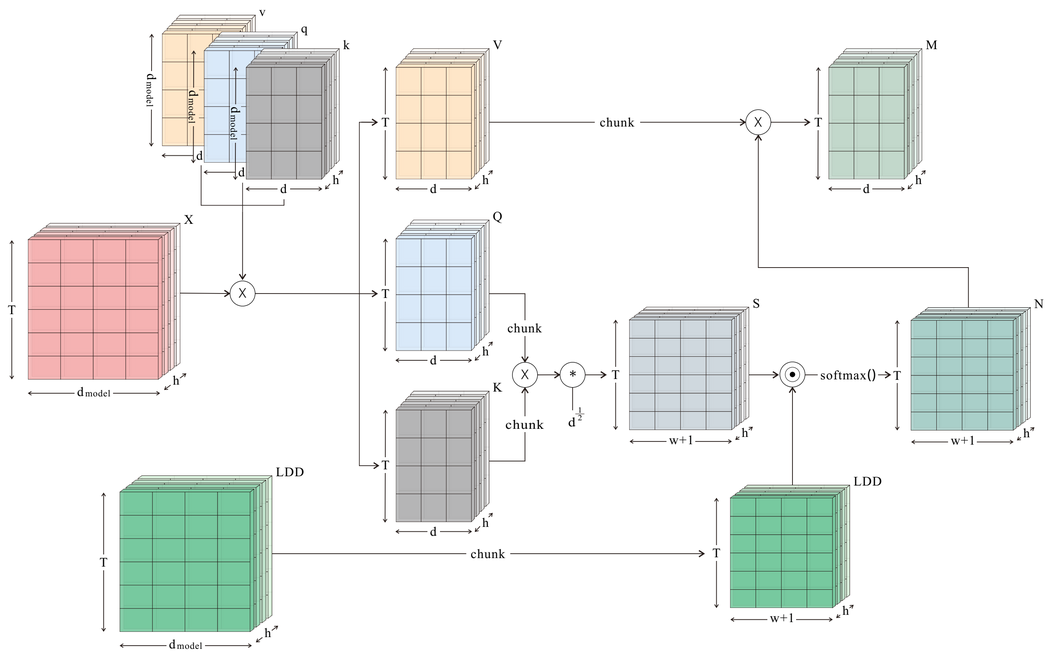

Figure 2. RSTformer architecture: we show a schematic diagram of incorporating LDD tensor into the attention layer of the model. Specifically, X is text embedding matrix, and LDD is incorporated with attention matrix S in the form of element-wise multiplication.

Next, the LDD tensor is injected into the attention tensor S via element-wise multiplication:

$$\mathrm{N} = \mathrm{S} \odot \mathrm{LDD} \quad (2)$$

Discourse-injected weights N are multiplied with the value matrix V to compute attention weights M for the present layer:

$$\mathrm{M} = \mathrm{N} \cdot \mathrm{V} \quad (3)$$

We infuse each attention layer with the LDD, and uniquely assign a distinct discourse matrix to each attention head, thereby enabling specialized concentration on diverse discourse labels.

Project Info

Paper

Main Results

Dataset	Model	Rouge-1 F1	Rouge-2 F1	Rouge-L F1	BERTscore	Meteor
BookSum Chapter	Full article (lower bound)	13.742	4.019	13.421	0.805	21.299
	Lead-3	17.683	2.747	16.708	0.812	9.815
	Lead-K	29.149	4.641	28.034	0.805	24.091
	Longformer (baseline)	33.636	9.626	32.611	0.846	27.160
	RSTformer (w/o relations)	33.604	10.149	32.631	0.850	26.811
	RSTformer (w/ relations)	34.019	**10.275**†‡	**32.870**	**0.853**†‡	**27.473**‡
	SOTA model (Kryscinski et al., 2022)	**37.510**	8.490	17.050	0.156	-
eLife	Full article (lower bound)	6.893	2.327	6.675	0.831	13.864
	Lead-3	16.266	3.634	15.088	0.832	7.163
	Lead-K	37.188	7.971	35.151	0.832	25.331
	Longformer (baseline)	46.778	13.318	44.317	**0.855**	27.921
	RSTformer (w/o relations)	46.862	14.008	44.458	**0.855**	27.685
	RSTformer (w/ relations)	**48.696**†‡	**14.843**†‡	**46.129**†‡	0.847	**29.526**†‡
	SOTA model (Goldsack et al., 2022)	46.570	11.650	43.700	-	-
Multi-LexSum	Full article (lower bound)	3.862	2.198	3.786	0.784	8.825
	Lead-3	16.135	6.387	15.421	0.770	9.538
	Lead-k	29.145	9.276	27.734	0.784	24.266
	Longformer (baseline)	45.751	21.272	43.131	0.865	33.282
	RSTformer (w/o relations)	46.424	22.730	43.978	0.867	33.808
	RSTformer (w/ relations)	46.421	22.888†‡	**43.979**	**0.867**‡	**33.941**
	SOTA model (Shen et al., 2022)	**53.730**	**27.320**	30.890	0.420	-

Table 1. Model performance. The bold numbers represent the best results with respect to the given test set. † and ‡ indicate statistical significance (p<0.05) against the baseline model via T-test and Kolmogorov-Smirnov test. Each result of the three distinct SOTA models is directly replicated from their original papers.

The results presented in Table 1 highlight the superior performance of RSTformer models compared to baseline models. These models exhibit improved **lexical choice** (reflected in Rouge & Meteor scores) and enhanced **semantic representation** (indicated by BERTscore). Notably, the RSTformer that incorporates discourse relation types outperforms both the variant without relation types and state-of-the-art models, implying the benefits of embedding more granular discourse information.

Analysis

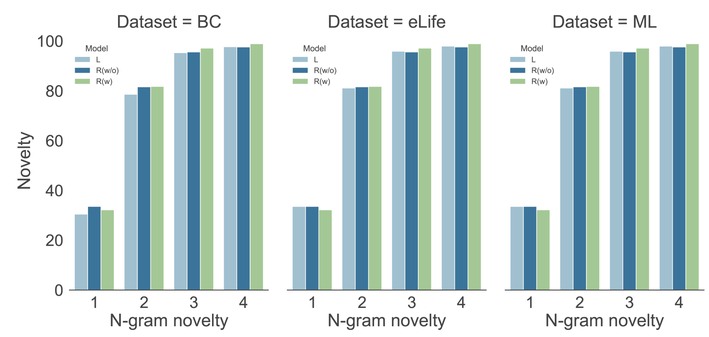

Figure 3. N-gram novelty. L = Longformer, R(w/o) = RSTformer(w/o relations), R(w) = RSTformer(w relations), BC = Booksum Chapter, ML = Multi-LexSum.

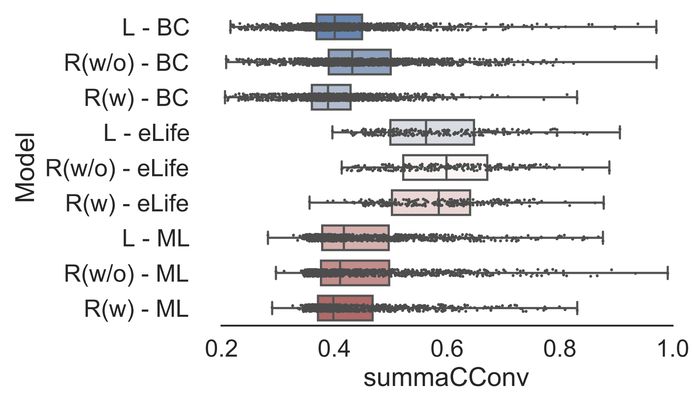

Figure 4. Consistency check

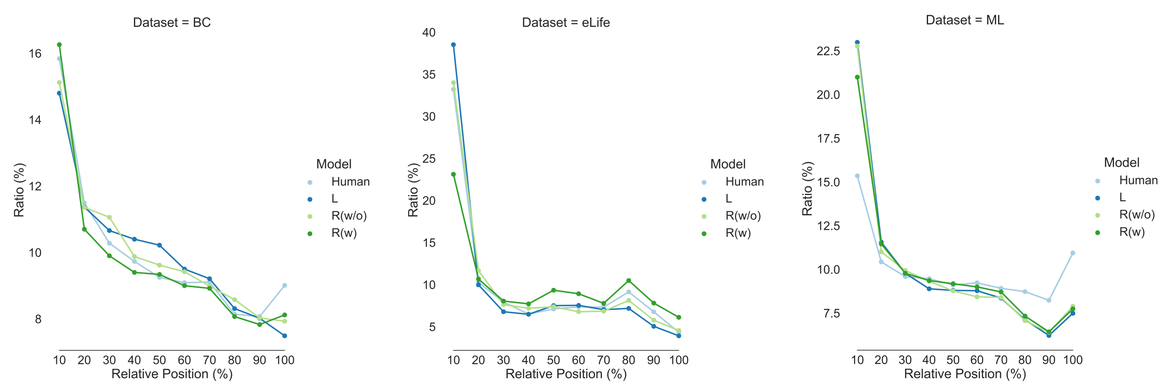

Figure 5. Sentence alignment distribution

Our in-depth analysis involving **sentence alignment**, **N-gram novelty**, and **inconsistency detection** further evidence for the effectiveness of our proposed RSTformer model. The model shows an increased ability to generate novel words, demonstrating enhanced abstractiveness. The model also outperforms the baseline in inconsistency checks, pointing to better factual consistency. Moreover, sentence alignment distributions show a close match with human summarizers, with a focus on content from the second half of the document, implying enhanced comprehensiveness and coherence.

Conclusion

In this study, we presented a novel supervised discourse-enhanced Longformer model, leveraging rhetorical structure as uncertainty distributions to enhance the local attention mechanism. Our experimental results convincingly demonstrate that this approach efficiently utilizes the discourse structure of source documents to bolster summary performance, exhibiting potential for broader applicability in other sequence-to-sequence natural language generation tasks.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 948878).