National Central University

Sequence to Sequence Text Generation: Text Simplification with Text-to-Text Transfer Transformer (T5)

Dicky P. Halim Robit Al Hazmi Andre Pratama Thien Nguyen

Group 5

Department of Computer Science and Information Engineering, National Central University

Background and Contribution Highlights

Text simplification, which is transforming complex text into simpler language while retaining its meaning, has several importance in NLP such as assisting readers that have reading difficulties, children, non-native speakers, and also pre-processing for NLP tasks: improves parsing, translation, summarization. However, this task has some challenges, such as balancing simplicity, fluency, and meaning. We tackled this challenge by leveraging Text-to-Text Transfer Transformer (TF) for text simplification task.

- We demonstrate T5's effectiveness in text simplification with robust performance on benchmarks.
- Highlight T5's unified framework for Seq2Seq text generation
- The importance of pretraining objectives and fine-tuning strategies is thoroughly evaluated.

Related Works

There are some recent advanced works on text generation tasks, one of them is DifFuSeq (Gong et al. 2022), which utilize Diffusion Model on Seq2Seq text generation tasks, including text simplification, highlighting its high computational cost but unique diffusion-based approach. There exists other diffusion-based approaches for text generation, Hoogeboom et al. (2021) introduce the multinomial diffusion for character-level text generation, the forward categorical noise is applied through the Markov transition matrix.

However, this approach is impractical and has high computational cost. More reliable approach could be Text-to-Text Transfer Transformer (T5) model which is a unified framework for handling a wide range of NLP tasks by casting them into a text-to-text format. Sheang et al. (2021) explores fine-tuning T5 for text simplification, incorporating control tokens to adjust outputs for different target audiences. The model achieved significant improvements over previous state-of-the-art methods.

Dataset: Wiki-Auto

The Wiki-Auto dataset is designed for training and evaluating text simplification models, using refined Wikipedia revisions. Unlike earlier noisy datasets, it ensures quality by employing a CRF aligner (Jiang et al., 2020) with fine-tuned BERT for semantic similarity and a paragraph alignment algorithm, leveraging the similar content order in parallel documents.

	Newsela		Wikipedia	
	Auto	Old	Auto	Old
# of article pairs	13k	7.9k	138k	65k
# of sent. pairs (train)	394k	94k	488k	298k
# of sent. pairs (dev)	43k	1.1k	2k	2k
# of sent. pairs (test)	44k	1k	359	359
avg. sent. len (complex)	25.4	25.8	26.6	25.2
avg. sent. len (simple)	13.8	15.7	18.7	18.5

Table 1. Dataset statistics for Newsela and Wikipedia.

Text-to-Text Transfer Transformer (T5)

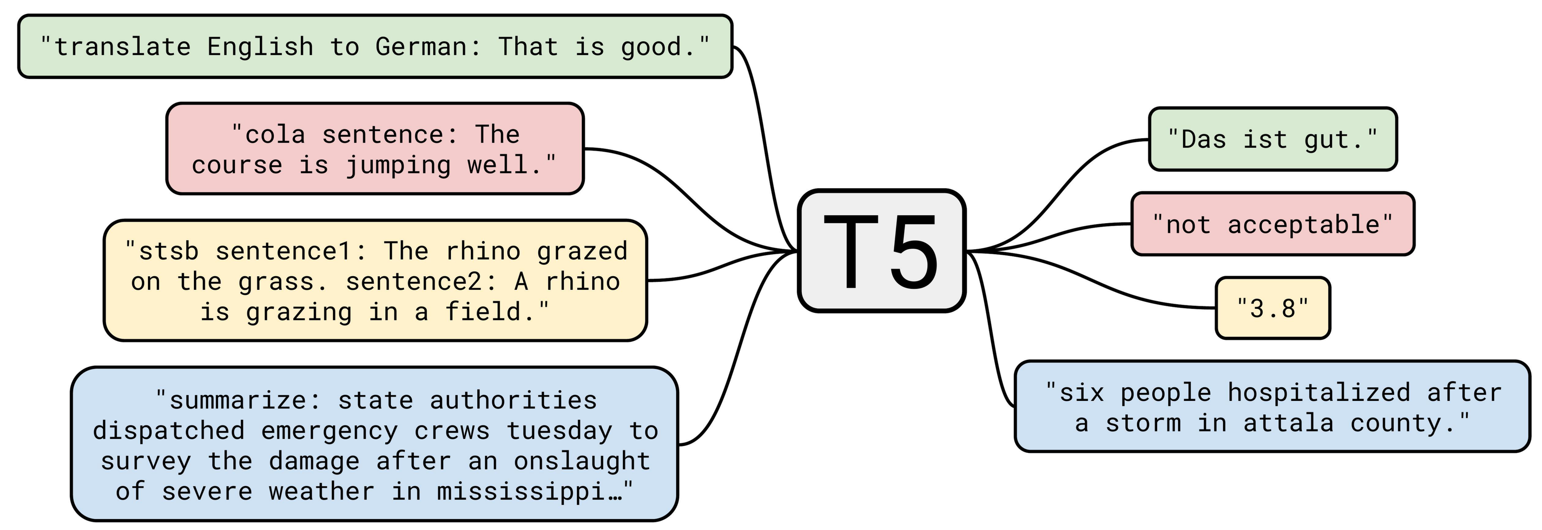

Figure 1. A diagram of the T5 framework

The basic idea underlying T5

T5 treats every text processing problem as a "text-to-text" problem, i.e. taking text as input and producing new text as output. T5 performance is evaluated on a wide variety of English-based NLP problems, including qna, text summarization, and sentiment classification (Raffel et al., 2020).

Key features of the T5 model

- **Text-to-Text Framework**: By framing every task as a text-to-text problem, T5 allows for more uniform and flexible approach to solve various NLP tasks.
- **Transfer Learning**: T5 leverages transfer learning, where the model is pre-trained on a massive corpus and fine-tuned on specific tasks.
- **Scalability**: T5 can be scaled up with larger models like T5–11B, which consists of 11 billion parameters.

Architecture of the T5 model

T5 model architecture follows its originally- proposed form (Vaswani et al., 2017). First, an input sequence of tokens is mapped to a sequence of embeddings, which is then passed into the encoder. The decoder is similar in structure to the encoder except that it includes a standard attention mechanism after each self-attention layer that attends to the output of the encoder.

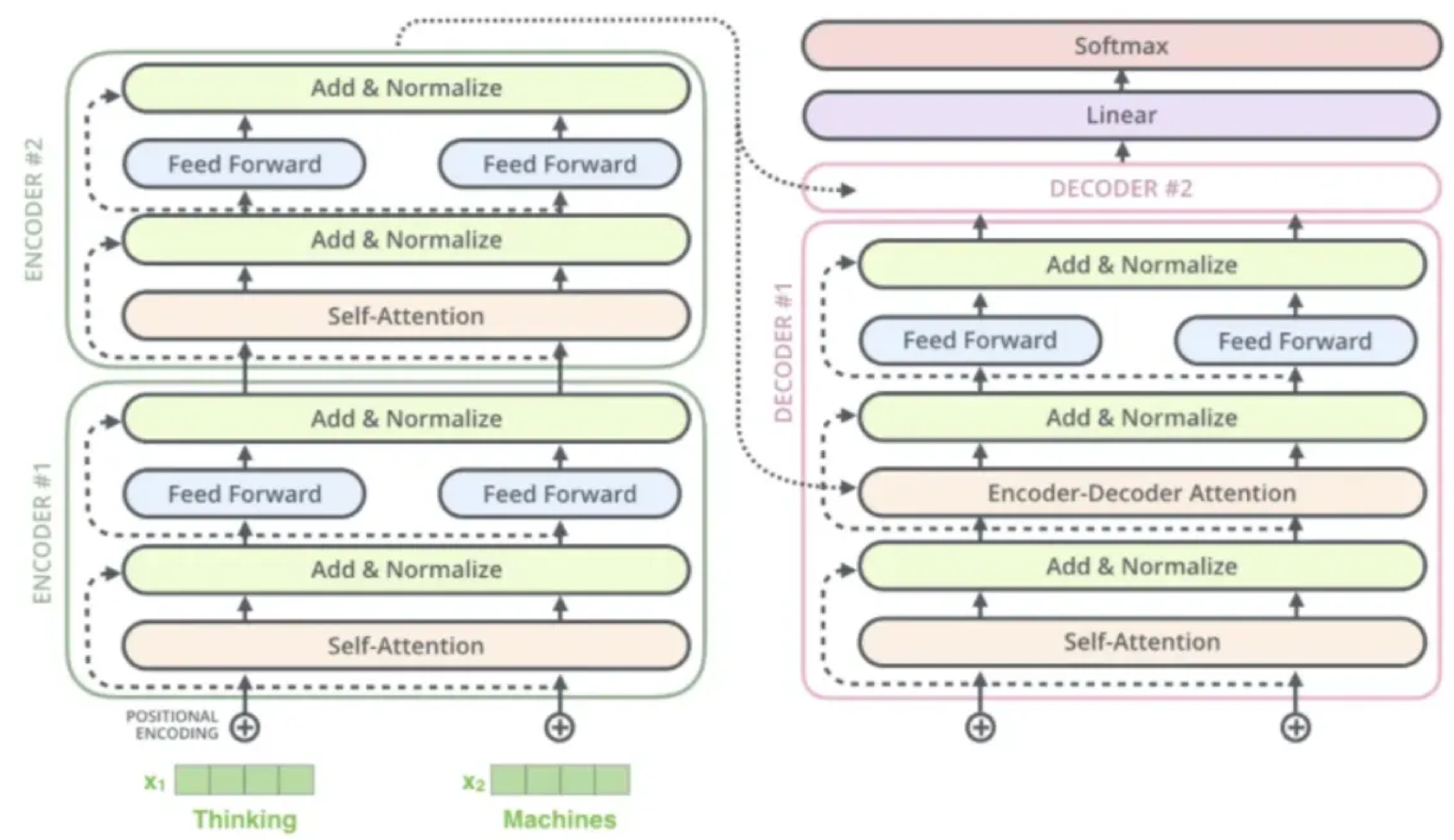

Figure 2. The Transformer - model architecture

Result for Text Simplification Task

Our result demonstrate that T5 achieves comparable or higher generation quality. We also provide examples to showcase our approach's ability to generate samples.

Tasks	Methods	BLEU↑	R-L↑	Score↑	Len
Text Simplification	GPT2-large FT	0.2693	0.5111	0.7882	15.4
	NAR-LevT	0.2052	0.4402	0.7254	8.31
	DIFFUSEQ	**0.3622**	0.584	0.8126	17.7
	T5-base FT (Ours)	0.3510	**0.5904**	**0.8237**	-

Table 2. Performance comparison across methods

Complex sentence: People can experience loneliness for many reasons, and many life events may cause it, such as a lack of friendship relations during childhood and adolescence, or the physical absence of meaningful people around a person.
Simplified: One cause of loneliness is a lack of friends during childhood and teenage years.

GPT2-large FT	NAR-LevT
• Loneliness can be caused by many things. • Loneliness can affect people in many ways. • Loneliness can be caused by many things.	• People may experience reashapphapphapphapphappabout life reasit. • People may experience reashapphapphapphapphappabout life reasit. • People may experience reashapphapphapphapphappabout life reasit.
DIFFUSEQ	**T5-base FT (Ours)**
• Many life events may cause of loneliness. • People can also be very experience loneliness for many reasons. • People can experience loneliness for many reasons, and many life events may cause it.	• People can experience loneliness for many reasons, and many life events may cause it. • People can experience loneliness for many reasons. • Many life events may cause loneliness.

Table 3. Simplification outputs for a complex sentence across models.

Conclusion

Our results show that T5 achieves comparable or superior performance in text simplification tasks compared to the other methods. Its unified framework and fine-tuning strategies highlight its effectiveness in generating fluent and meaningful simplified text, showcasing T5's potential for broader NLP applications.

References

[1] Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and LingPeng Kong. Diffuseq: Sequence to sequence text generation with diffusion models. *arXiv preprint arXiv:2210.08933*, 2022.

[2] Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. Neural crf model for sentence alignment in text simplification. *arXiv preprint arXiv:2005.02324*, 2020.

[3] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67, 2020.

[4] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.