Lexical Simplification via Single-Word Generation

Jipeng QIANG, Yang LI, Yun LI, Yunhao YUAN, Yi ZHU

Frontiers of Computer Science, DOI: 10.1007/s11704-023-2744-2

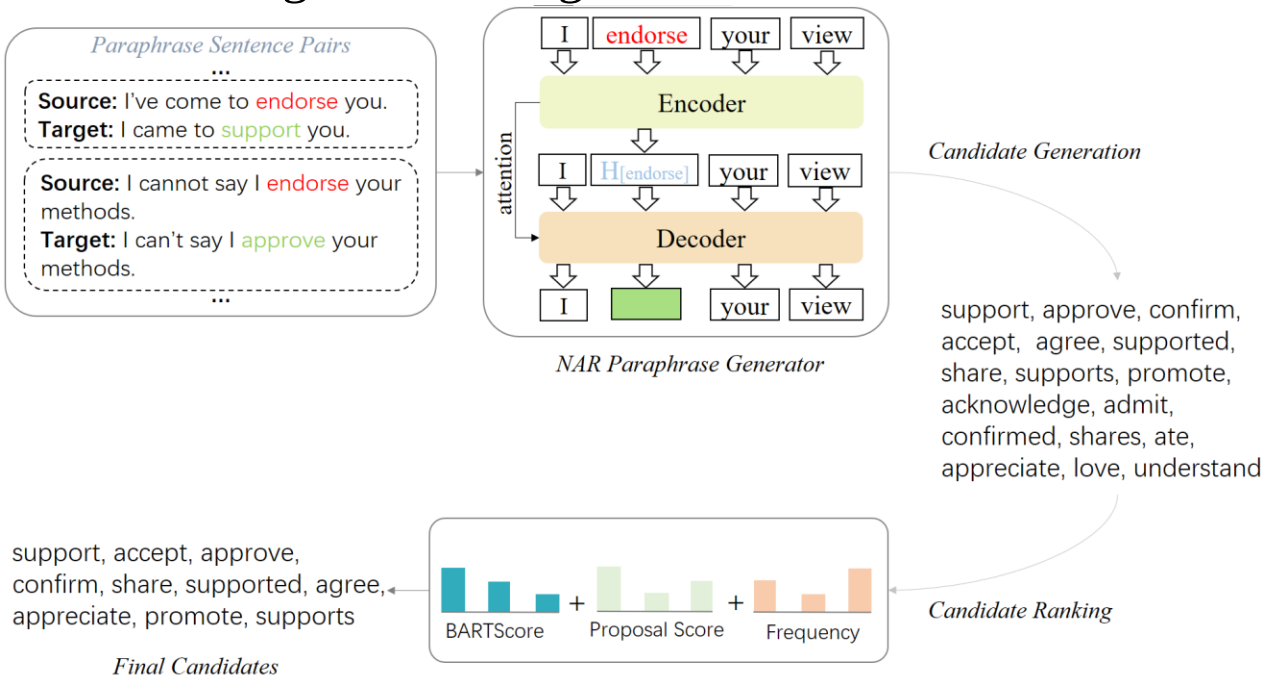
Problems & Ideas

Problems of BERT-based LS method :

- BERT as a self-supervised pretrained model that is trained with the goal of recovering the destroyed original text, does not significantly learn the word substitution operation.
- Masking the complex word will impair the semantic information of complex word, resulting in failing to preserve the sentence's meaning.

Ideas:

- The generation of candidate words by the paraphrase generation model can well preserve the original meaning of the sentence.



The framework of PaGeLS

Experimental Result & Conclusions:

Experimental Result :

Table 1. Evaluation result of simplification candidate generation..

	LexMTurk			BenchLS			NNSeval		
	PRE	REC	F1	PRE	REC	F1	PRE	REC	F1
Glavaš	0.151	0.122	0.135	0.142	0.191	0.163	0.105	0.141	0.121
Embed	0.177	0.140	0.156	0.180	0.252	0.210	0.118	0.161	0.136
BERT-LS	0.287	0.223	0.251	0.231	0.314	0.265	0.185	0.246	0.211
LSBert	0.306	0.238	0.268	0.244	0.331	0.281	0.194	0.260	0.222
PaGeLS	**0.371**	**0.297**	**0.330**	**0.287**	**0.382**	**0.327**	**0.221**	**0.284**	**0.249**
w/o Ranking	0.337	0.270	0.300	0.264	0.352	0.301	0.206	0.266	0.232

Table 2. Full pipeline evaluation results using Precision and Accuracy.

	LexMTurk		BenchLS		NNSeval	
	PR	ACC	PR	ACC	PR	ACC
Glavaš	0.710	0.682	0.480	0.252	0.456	0.197
Embed	0.578	0.396	0.423	0.423	0.297	0.297
BERT-LS	0.770	0.770	0.604	0.604	0.420	0.420
LSBert	0.864	0.792	0.697	0.616	0.526	0.436
PaGeLS	**0.874**	**0.874**	**0.723**	**0.723**	**0.527**	**0.527**

Conclusions:

- Proposes to generate candidate words based on a paraphrase generation model without changing the meaning of the sentence.
- The text generation evaluation metric BARTScore is proposed to evaluate the suitability of candidate words.
- PeGeLS achieves state-of-the-art results compared with the best BERT-base method.