

T5

- ▶ Pre-training: similar denoising scheme to BART (they were released within a week of each other in fall 2019)
- ▶ Input: text with gaps. Output: a series of phrases to fill those gaps.

Original text

Thank you ~~for inviting~~ me to your party ~~last~~ week.

Inputs

Thank you <X> me to your party <Y> week.

Targets

<X> for inviting <Y> last <Z>

T5

Number of tokens	Repeats	GLUE	CNNNDM	EnDe	EnFr	EnRo
★ Full dataset	0	83.28	19.24	26.98	39.82	27.65
2^{29}	64	82.87	19.19	26.83	39.74	27.63
2^{27}	256	82.62	19.20	27.02	39.71	27.33
2^{25}	1,024	79.55	18.57	26.38	39.56	26.80
2^{23}	4,096	76.34	18.33	26.37	38.84	25.81

summarization

machine translation

- ▶ Colossal Cleaned Common Crawl: 750 GB of text
- ▶ T5 was one of the first papers to heavily focus on the scale and quality of pre-training data
- ▶ Another key focus: multitasking between generation tasks

Successes of T5: UnifiedQA

- ▶ How can we handle a task like QA by framing it as a seq2seq problem?

Dataset	SQuAD 1.1
Input	At what speed did the turbine operate? \n (Nikola_Tesla) On his 50th birthday in 1906, Tesla demonstrated his 200 horsepower (150 kilowatts) 16,000 rpm bladeless turbine. ...
Output	16,000 rpm

- ▶ Format: *Question* \n *Passage* → *Answer*
 encoder decoder

Successes of T5: UnifiedQA

	Dataset	NarrativeQA
AB	Input	What does a drink from narcissus's spring cause the drinker to do? \n Mercury has awakened Echo, who weeps for Narcissus, and states that a drink from Narcissus's spring causes the drinkers to ``Grow dotingly enamored of themselves.'' ...
	Output	fall in love with themselves

Abstractive question, requires generating *free-form answer*

- ▶ Past work: different architectures for every QA formulation. (Span selection, answer generation, multiple choice, ...)
- ▶ Now: one 11B parameter T5 model

Successes of T5: UnifiedQA

Multiple choice

MC	Dataset	MCTest
	Input	Who was Billy? \n (A) The skinny kid (B) A teacher (C) A little kid (D) The big kid \n Billy was like a king on the school yard. A king without a queen. He was the biggest kid in our grade, so he made all the rules during recess. ...
	Output	The big kid
YN	Dataset	BoolQ
	Input	Was America the first country to have a president? \n (President) The first usage of the word president to denote the highest official in a government was during the Commonwealth of England ...
	Output	no

Yes/no

- ▶ Past work: different architectures for every QA formulation. (Span selection, answer generation, multiple choice, ...)
- ▶ Now: one 11B parameter T5 model
- ▶ Gives strong results across all these QA tasks when fine-tuned on them together