

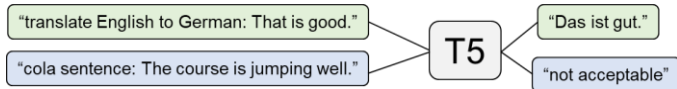
Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (T5)

JMLR 2020

Tag: Text-to-Text, Transfer learning, Transformer

0. Summary

- T5: Text-to-Text Transfer Transformer



- Text-to-Text: input-output format이 모두 text, 동일한 loss function과 decoding 과정으로 사용 가능한 single model
- Transfer learning: 특정 task를 학습한 모델을 다른 task 수행에 재사용
- Transformer: Transformer의 구조를 사용

1. Setup

> Pre-training

- C4 Dataset 사용 (The Colossal Clean Crawled Corpus)
- Common Crawl data를 수집 후 전처리 진행
- 대부분 영어로 되어있음, 750GB 정도 수집함

> Fine-Tuning

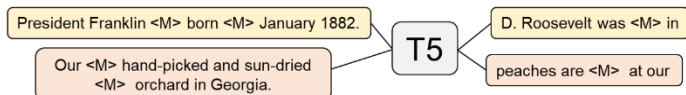
- Text classification, Abstract Summarization, Question answering, Translation에 대해 진행

> Input & Output format

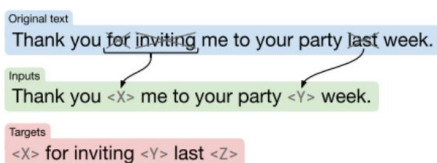
- 모든 문제를 Text-to-Text로 바꾸어서 처리함
- Regression task의 경우 classification 문제로 변경
→ 1~5사이의 값 → "1.0", "1.2", ..., "4.8", "5.0"
- Classification Task의 경우 다른 라벨이 생성되는 경우 틀린 것으로 처리

2. Training

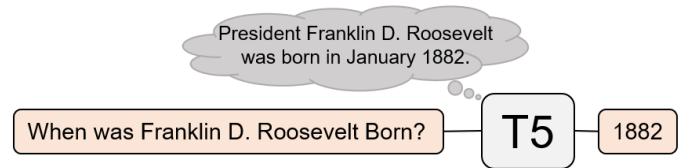
> Pre-training



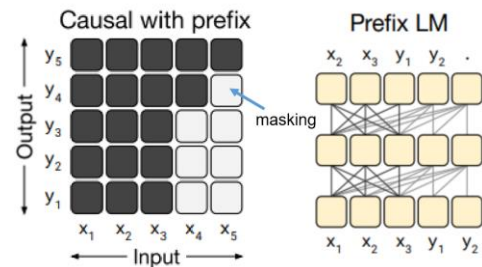
- C4 dataset사용
- Input의 15%의 token에 Masking
- 연속되어 Masking 될 경우, 하나의 sentinel token으로 처리 (Denoising Corrupted Span)



> Fine-Tuning



3. Architecture

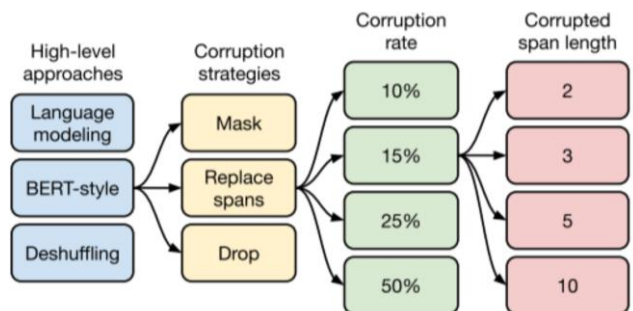


* X: translate English to German: That is good. target:

* Y: Das ist gut

- Model은 전체 Input에 대해서 계속 볼 수 있으며 접두사를 학습함
- "target: "가 BERT에서 [CLS]과 비슷한 역할을 함

4. Experiments



- 우선순위를 정하여 실험 진행
- BERT-style(MLM) → Replace spans(좌측 2. pre-training 예시) → Mask 비율 15% → Baseline(랜덤 Masking - 연속된 span을 하나로) 일 때가 BEST
- Pre-training data는 C4로 했을 때가 가장 성능이 좋음
- 다른 Data 의 경우 성능이 더 높은 것도 있었으나, 데이터셋의 크기가 작아서 overfitting 이 발생

끝!!

질문 있으시면 정민지에게 DM 주세요!!