

RoBERTa

practice 최신트랜드 논문 · 2024. 10. 12. 16:48

Q. RoBERTa (A Robustly Optimized BERT Pretraining Approach) - 일반화가 잘되었다는 모델

A. Bert 강화버전

학습데이터 늘리고, 학습시간늘림.

NSP빼고 Dynamic masking써서 기존 Bert보다 많이 올렸음

XLNet 다음에 나온 논문

일반화가 잘되었다는 모델

Bert를 잘 훈련시키겠다.

hyperparameter를 바꿔가며 Bert를 더 훈련시킴

Bert가 학습이 충분히 안되었다.

기존 Bert에 비해서 10배정도 Data를 더 많이 썼다.

Training을 더 오래돌림(data가 많아서)

Batch size 를 키움

Bert = MLM(masked language model) + NSP(next sentence prediction)

인데

RoBERTa = MLM(masked language model) 만 씀

Dynamic masking써서 masking을 중간에 바꿔서 썼음

Model	data	bsz	steps	SQuAD (v1.1/2.0)	MNLI-m	SST-2
RoBERTa						
with BOOKS + WIKI	16GB	8K	100K	93.6/87.3	89.0	95.3
+ additional data (§3.2)	160GB	8K	100K	94.0/87.7	89.3	95.6
+ pretrain longer	160GB	8K	300K	94.4/88.7	90.0	96.1
+ pretrain even longer	160GB	8K	500K	94.6/89.4	90.2	96.4
BERT _{LARGE}						
with BOOKS + WIKI	13GB	256	1M	90.9/81.8	86.6	93.7
XLNet _{LARGE}						
with BOOKS + WIKI	13GB	256	1M	94.0/87.8	88.4	94.4
+ additional data	126GB	2K	500K	94.5/88.8	89.8	95.6

♡ 1



...

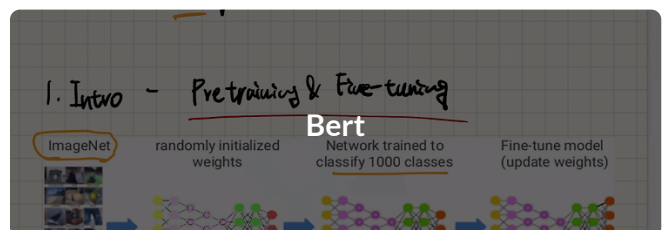
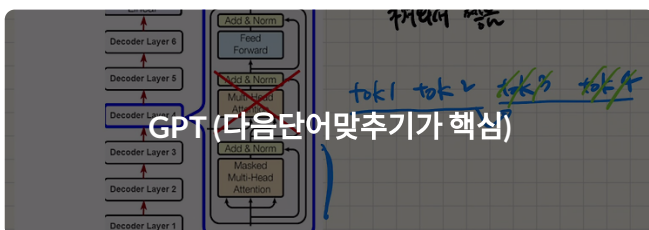
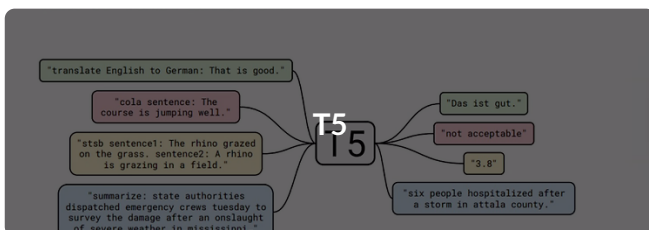
구독하기

'practice 최신헤랜드 논문' 카테고리의 다른 글

T5 (1)	2024.10.12
Bert 이후 모델 (Transfomer-XL, GPT2, XLNet). (2)	2024.10.12
GPT (다음단어맞추기가 핵심). (1)	2024.10.07
Bert (0)	2024.10.04

관련글

[관련글 더보기](#)



자연어(NLP)

네이쳐2024 님의 블로그입니다.



구독하기 +

댓글 0



이름

비밀번호

내용을 입력하세요.



등록