Bert

practice 최신트랜드 논문 · 2024. 10. 4. 15:49

Q: BERT 논문의 주요 기여와 모델의 작동 방식
A: 문맥을 양방향으로 이해할 수 있는 모델

작동 방식
• Pretraining
1. Masked LM: 마스킹 된 단어를 맞춤으로써 문맥 이해
2. NSP: 두 문장이 이어진 문장인지 판별함으로써 자연어 이해
• Finetuing

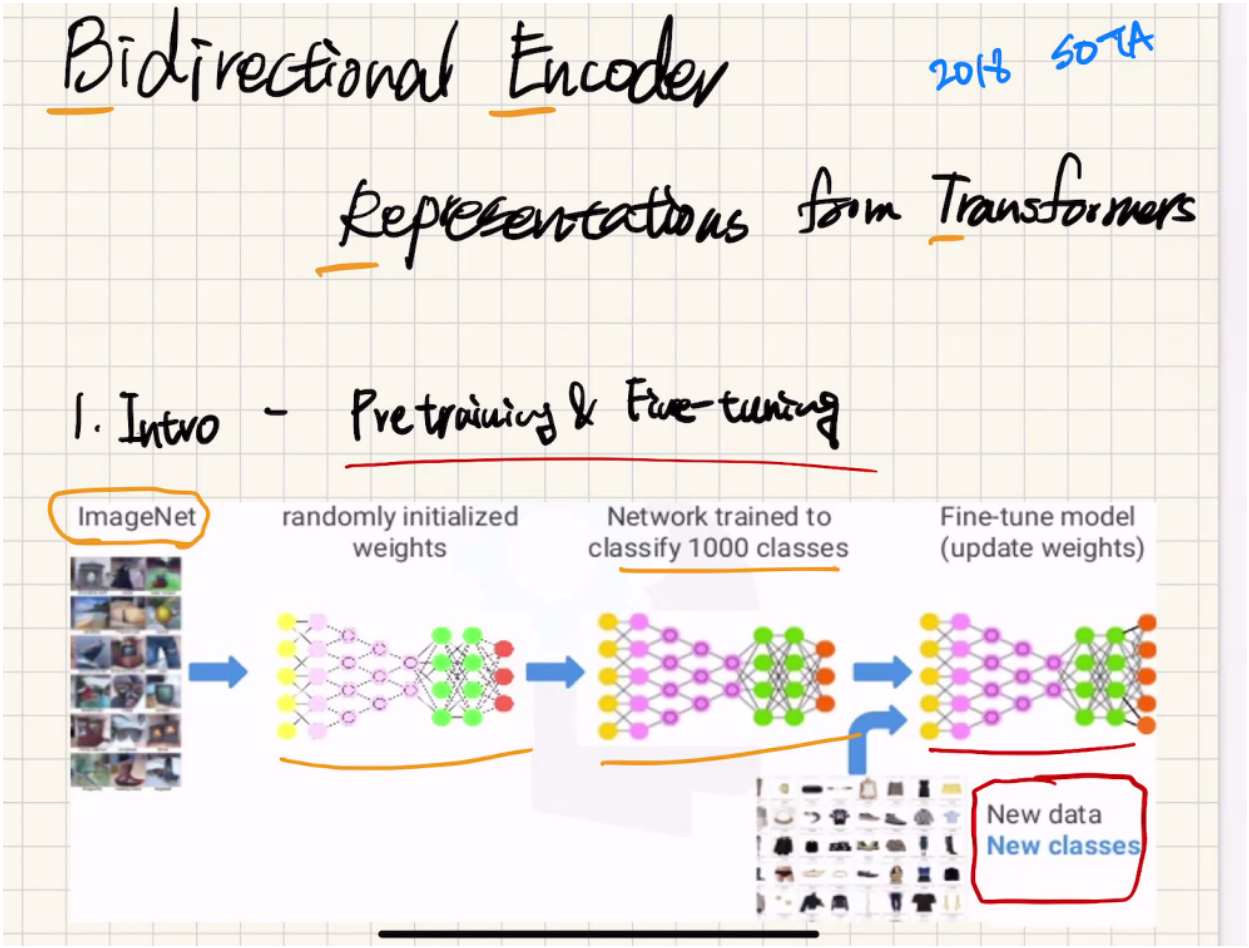
1. Bert (내부구조는 transformer의 Encoder부분만 쌓음)
Bidirectional Encoder Representation from Transformers

2018 SOTA(State of the art - 그때 당시 최고 성능이 제일 좋은 모델)

기존에 양방향 모델이 있었는데(Bi RNN in paper ELMO) 한쪽(앞에서 뒤쪽)으로 갔다가, 다른 한쪽(뒤쪽에서 앞쪽)으로 가서 각각 한쪽으로 간 정보를 이어붙인 상태여서 진정한 Bidirectional이 아니었다.
GPT1은 한쪽으로만 정보가 가는 모델이었다.
그래서 양방향에서 맥락을 이해하는 모델이 필요했다(앞쪽을 봐서 뒤쪽유추, 뒤쪽에서 앞쪽유추, 양쪽을 보고 가운데 유추가능모델링 필요)
그래서 랜덤하게 단어를 마스크하고 이것을 맞추는 작업을 pretrained시켰다.

1. Introduction - pretraining & Finetuning

pretraining과 finetuning 개념을 이미지쪽에서 갖고와서 처음 가져온 모델중 자연어쪽으로 잘 가져온 모델중하나

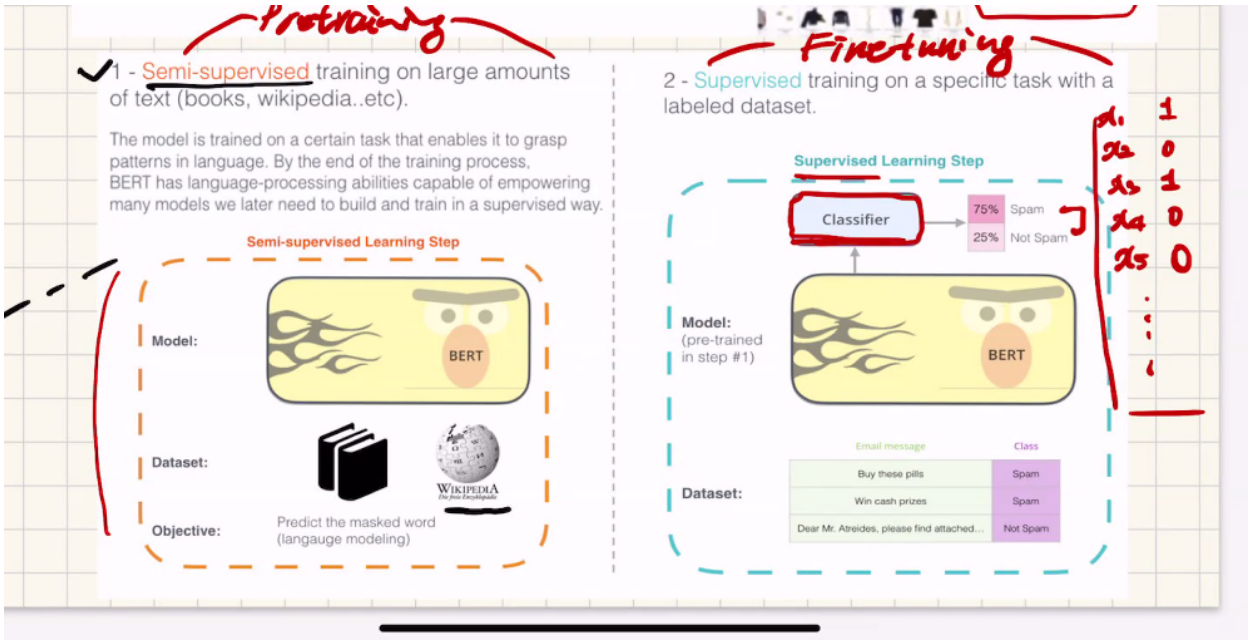


아래그림 - pretraining 과정과 finetuning 과정

pretraining 과정과 finetuning 과정으로 진행되는데,

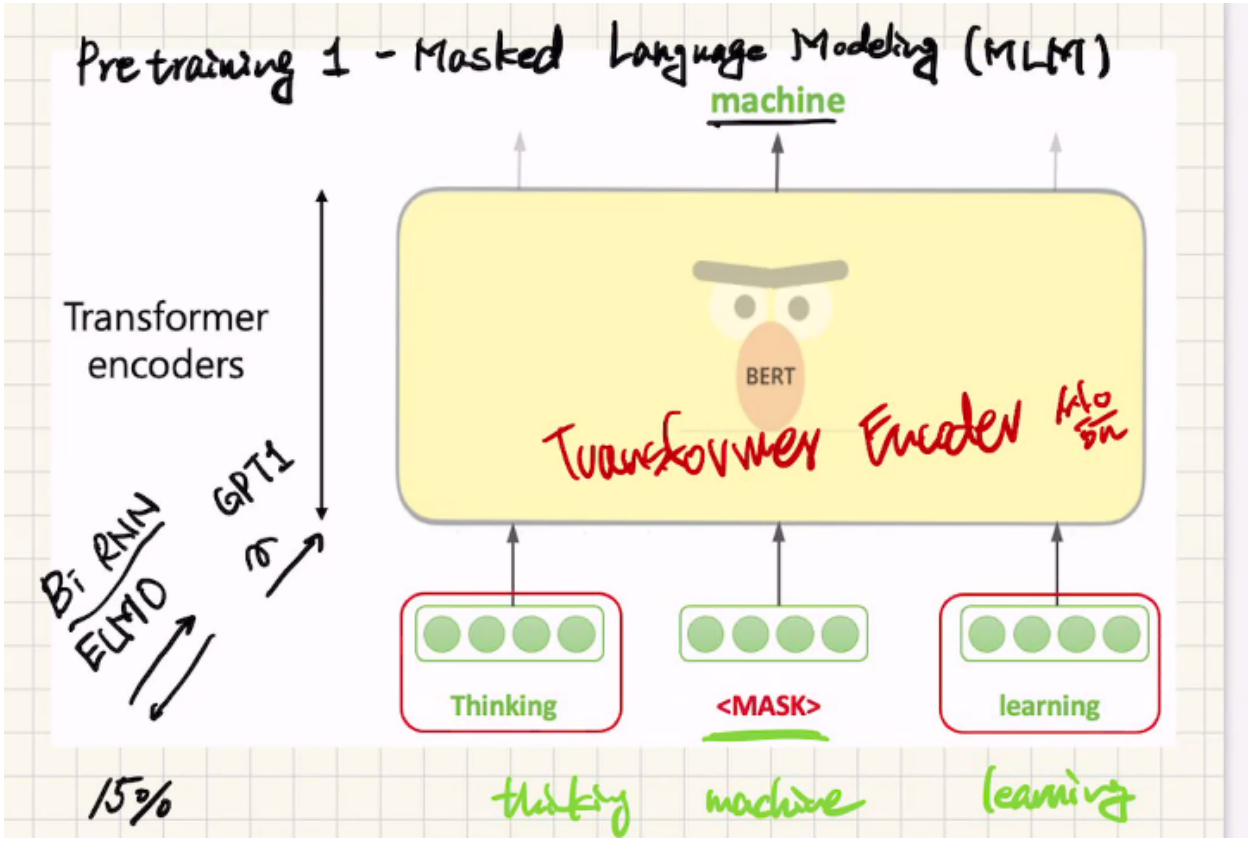
pretraining 과정은 semi supervised training으로 가운데단어 맞추기를 학습한후 데이터에 있는 단어로 test한후,

finetuning 과정은 이 pretraining한 모델을 specific task (예, spam인지 아닌지 분류) 로 정답이 있는 데이터로 분류하는 supervised과정이다.



아래그림 - pretraining

1 - Masked Language Modeling(MLM)

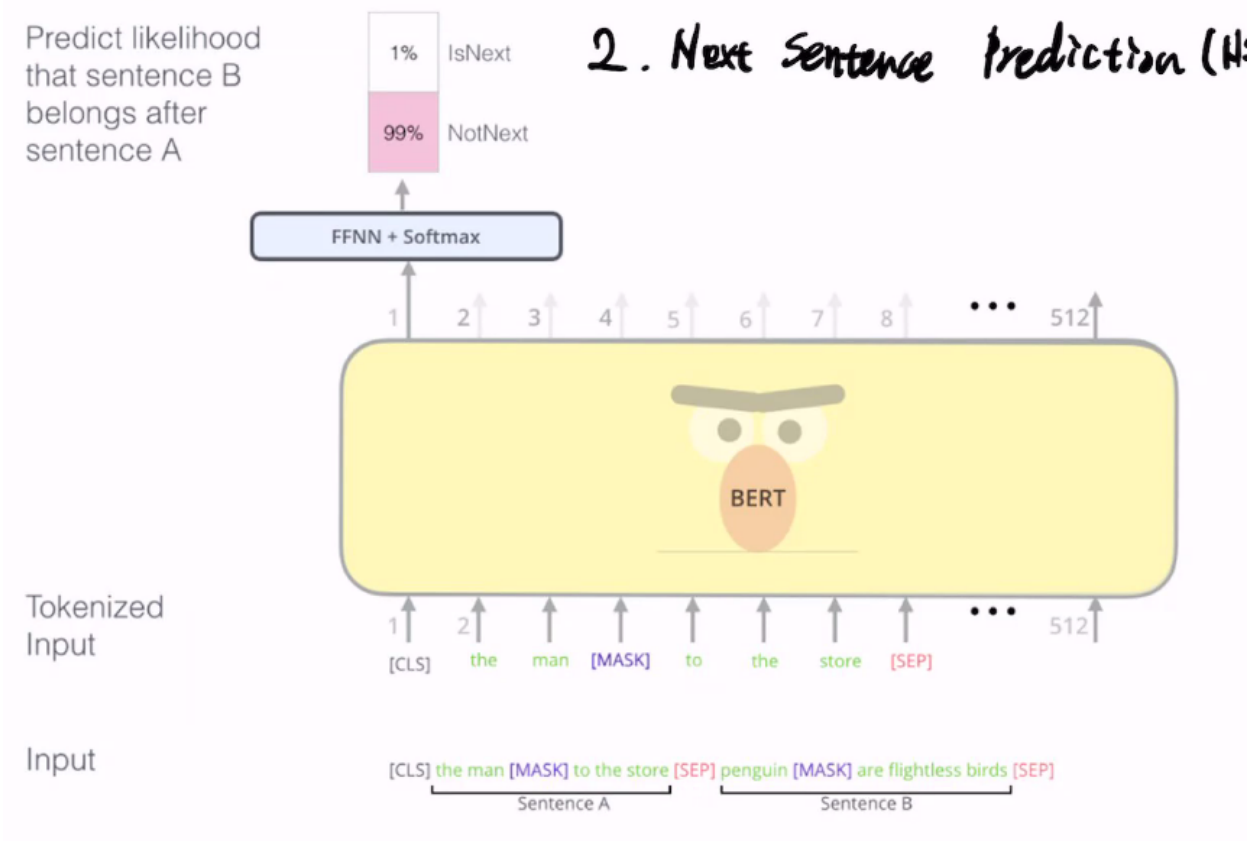


2. Next Sentence Prediction (NSP)

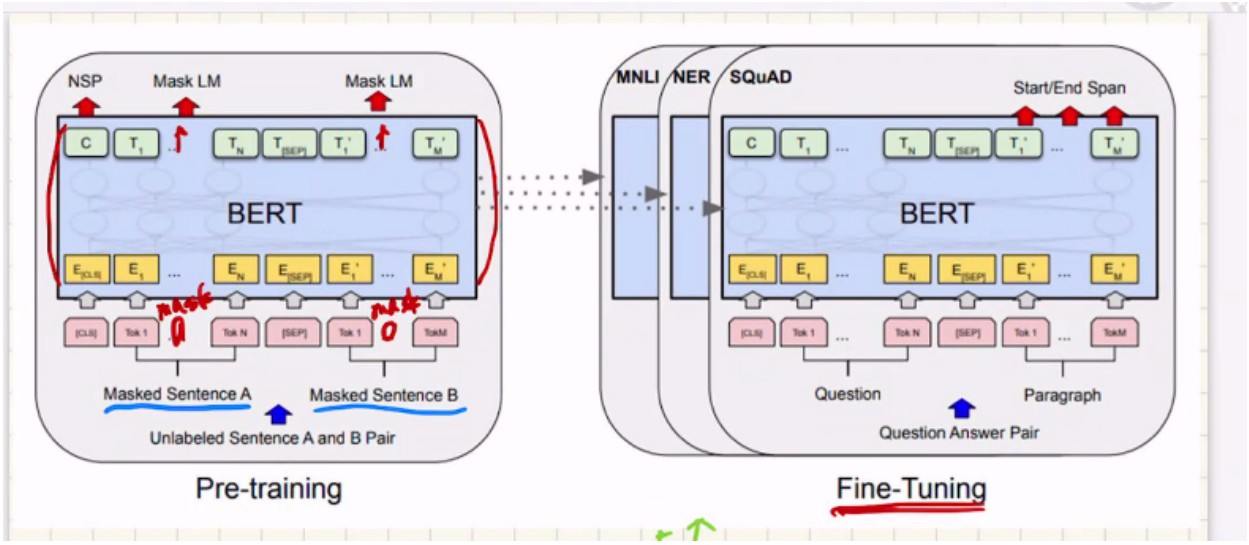
두문장을 붙여놓고 실제로 붙어있는건지 데이터에서 임의로 갖고와서 2개를 붙여놓은지를 확인(문장간의 관계를 이해하기 위함)
두문장을 가져와서 두개가 연관됬는지를 확인

Bert pretrain 목적 - 자연어를 잘 이해하는 모델 만드는것

Next Sentence Prediction (NSP) 목적- 문장들에 순서대로 index를 붙여서 두개를 가져와서 연관성있는지 비교



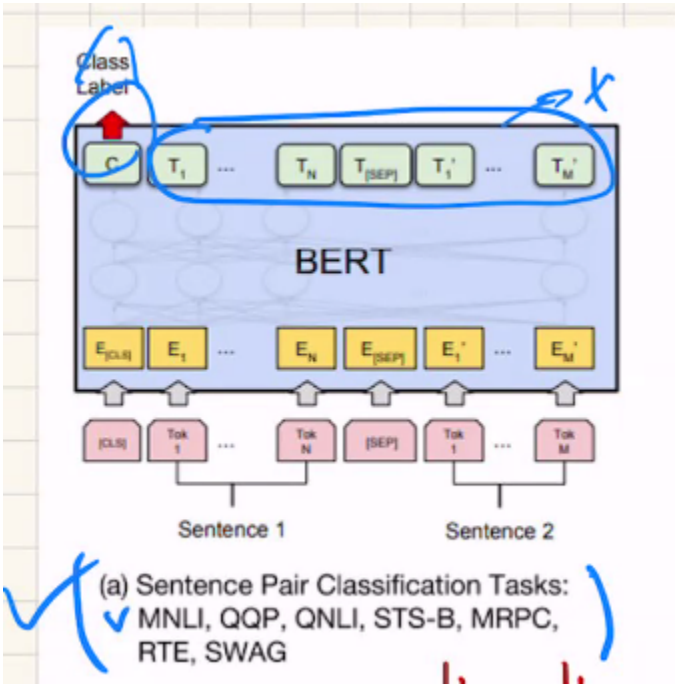
아래그림)
(왼쪽그림) pretrain: masked sentence 두개를 입력해서 정답을 맞추는 과정 -> next sentence prediction으로 cls를 이용해서 원래 두개문장이 이어진 문장인지 맞추는과정이랑, masked token을 을 맞추는 과정
(오른쪽그림) finetuning - pretrain모델을 가져와서 조금씩만 변형해서 다양한 작업을 가능하게 한다(예: Sentence Pair Classsification Tasks, single sentence classification task, Question Answering Task, single sentence tagging task)



아래그림)

Sentence Pair Classsification Tasks

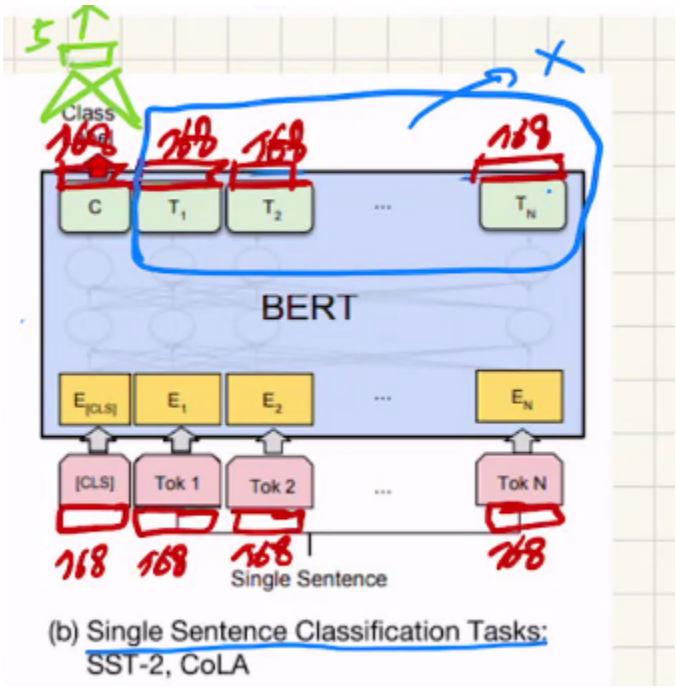
sentence 2개를 입력해서 맨앞만 남기고 나머지는 버리고 이 둘의 관계성을 분류한다.



아래그림)

single sentence classification task

각 token은 768 벡터크기인데 bert모델을 통과해서 맨앞에만 원하는 리니어 사이즈(예 크기 5)로 출력한다.



아래그림)

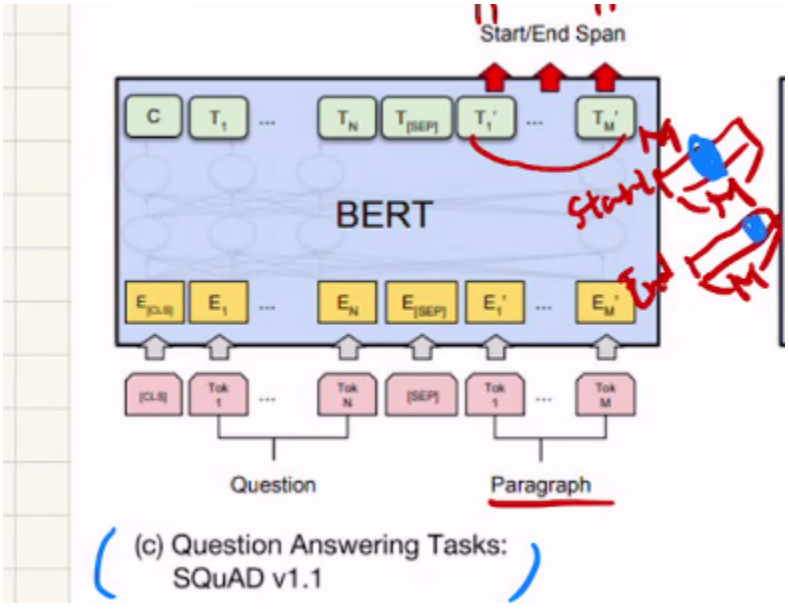
Question Answering Task

지문과 질문을 주고 지문내에서 답을 찾는 task

질문의 답이 지문에서 paragraph가 어디에서 시작하고 어디서 끝나는지 찾는다.그래서 시작과 끝의 logit을 뽑고

각각의 위치에서 숫자를 2개씩 뽑아서 시작은 시작기리 모으고 끝은 끝끼리 모으면 크기 m의 벡터를 모을수 있다.

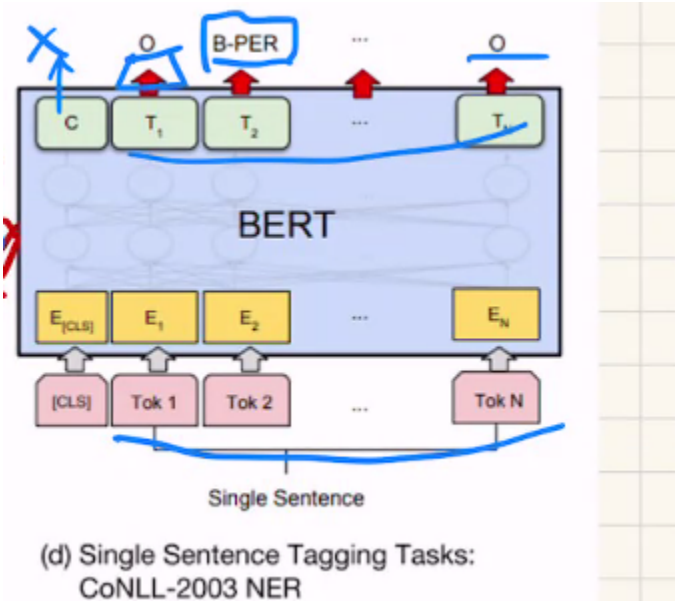
거기서 숫자가 제일 높은부분이 시작과 끝이다.



아래그림)

single sentence tagging task

NER분류로 각 token이 person 인지 organization인지등으로 분류



(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

GPT1: BookCorpus(800M)

BERT: BookCorpus + wikipedia(2500M)

아래그림: Bert 성능이 제일 높음

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**92.7**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**82.1**

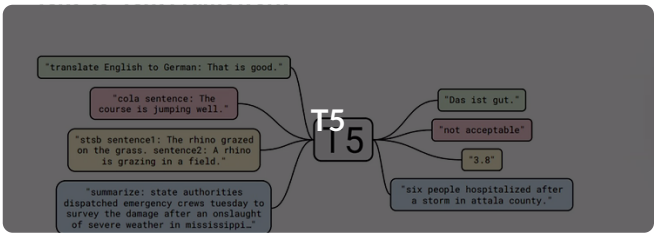
Table 1: GLUE Test results, scored by the evaluation server (https://gluebenchmark.com/leaderboard). The number below each task denotes the number of training examples. The "Average" column is slightly different than the official GLUE score, since we exclude the problematic WNLI set.[8] BERT and OpenAI GPT are single-model, single task. F1 scores are reported for QQP and MRPC, Spearman correlations are reported for STS-B, and accuracy scores are reported for the other tasks. We exclude entries that use BERT as one of their components.

'practice_최신트랜드 논문' 카테고리의 다른 글

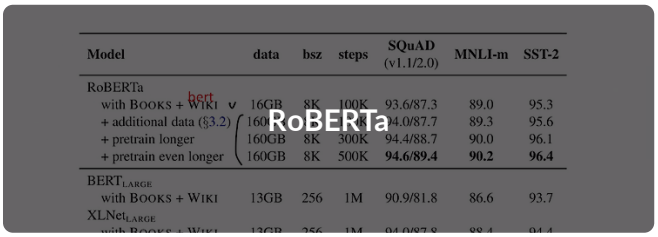
T5 (1) 2024.10.12

관련글

관련글 더보기

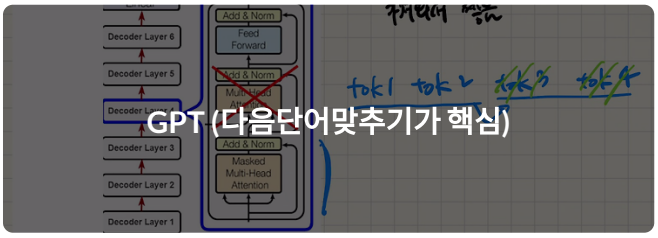


T5



RoBERTa

Bert 이후 모델 (Transfomer-XL, GPT2, XLNet)



GPT (다음단어맞추기가 핵심)

자연어(NLP)
네이쳐2024 님의 블로그입니다.
구독하기 +

댓글 0

이름

비밀번호

내용을 입력하세요.