

BERT: Deep-Bidirectional Transformers for Language Understanding

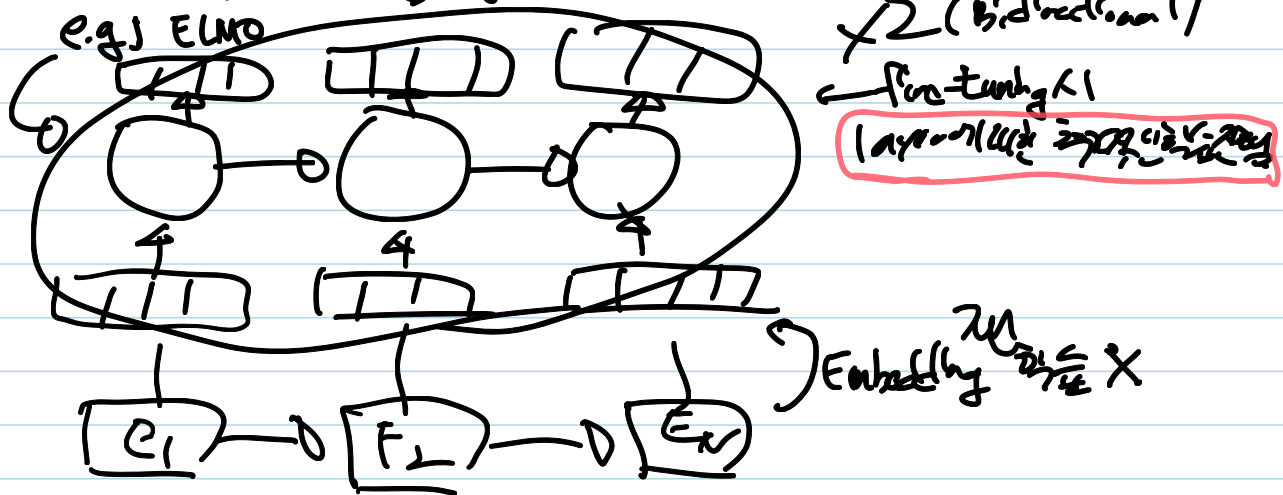
Abstract: 문맥에 있는 정보와 문맥에 있는 정보를 함께 고려하여 Labeling 되어 있지 않은 text를 통해 pre-train 진행

⇒ 결과적으로 마지막 output layer를 fine-tuning 해서
요약, 분류, 등 다양한 용도로 사용 가능

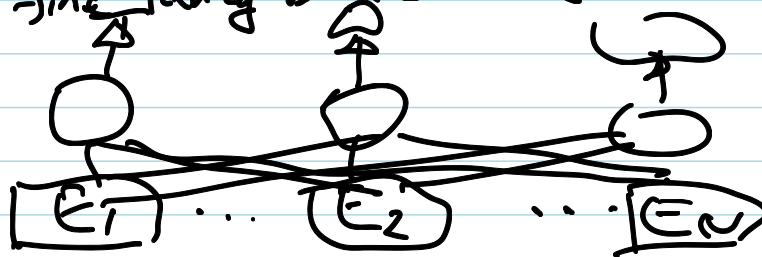
Intro: Preliminaries: task based learning VS fine-tuning based

pre-trained의 language representation을 down-stream tasks에 적용하는 2가지 방법

① task-based: 일대일 (task에 맞는 representation을 학습)



② fine-tuning based: e.g.) GPT



Intro: Main Idea

Contribution ①: 양방향 정보를 통한 language representation에 대한 향상

② 특정 task에 맞는 구조를 위해. how? 이 학습한 정보를
공유

Bert의 주요한 두가지 구조

- ① Pre-training (Self Supervision)
- ② Fine-tuning (Supervision)

① Model Architecture :

: 일반적인 Transformer 형태로 구성되어 있음.

② I/O Representation

- (Question, Answer) pair을 공부함
- Wordpiece Embedding 사용

③ Pre-training Bert

Task 1: Masked language model

현제 language model의 출력은 LOR, ROL의 일련성으로 출력되는

→ Wordpiece embedding의 약 15%를 마스킹하여 양방향으로 보충할 수 있음

하지만, 실제로 Mask token이 Fine-tuning 단계에서 등장 X

∴ 토큰 자체의 약 15%를 마스킹하여 정해진 토큰을 입력하여 Masking 가능

Task 2: Next Sentence Prediction

'두 문장 간의 관계는?'

(B는 A를 10가지로 나누고, 50%는 연결함)

④ Fine-tuning Bert

- 신신리 (입력 - 출력) pair을 사용해서 Fine-tuning 진행.

⑤ 실험결과

→ 다양한 Bert에 대해서 성능 향상

fine-tuning 기술로 인해 다양한 환경에서 적용 가능.