

AI504 23강 정리

- GPT-1
 - Generative Pre-Training -> GPT
 - Forward Language Modeling만 사용 (Transformer의 디코더 부분)
 - BooksCorpus로 pre-train(perplexity 18.4)
 - 구체적인 downstream architecture 없음, 그저 하나의 linear layer
 - ◆ Classification, Entailment, Similarity, Multiple Choice
- BERT
 - Bidirectional Encoder Representations
 - ELMo의 Bidirectionality + GPT의 Powerful Transformer
 - 구조
 - ◆ Token Embedding + Position Embedding + Segment Embedding
 - ◆ CLS: 입력 전체를 represent하는 special token
 - Pre-training
 - ◆ Masked Language Modeling (a.k.a Cloze Task)
 - 토큰 들 중 15%를 masking 하고 그것을 identify 하는 것
 - ◆ Next sentence prediction (NSP)
 - 두 문장이 실제로 neighbor면 CLS=1, 아니면 0
 - 일부 downstream task에 도움됨
 - ◆ BooksCorpus, English Wikipedia로 pre-train(MLM perplexity ~4)
 - Downstream task에 따라 다른 입출력 필요(+fine-tuning)
- GPT-2
 - GPT-1의 10배
 - Trained on WebText(perplexity 10~11)
 - ◆ 3 karma 이상의 reddit post
 - ◆ Wikipedia 제외 (benchmark dataset이 Wikipedia에 의존)

- Zero-shot task = Fine-tuning 없이 downstream task
 - ◆ 어떤 downstream task를 할지 알려줘야 함
 - Downstream task 몇 개를 예시로 넣고 그 다음에 sampling 함
 - ◆ 상당히 좋은 결과를 보임
- GPT-3
 - GPT-2의 100배
 - Zero-shot, One-shot, Few-shot
 - ◆ Task를 묘사한 후 예시를 몇 개 보여줄 것인가