<https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf>

**모델의 아키텍쳐는 어떠한가?**

GPT는 transformer의 디코더 아키텍처를 활용한다. Encoder은 context vector로 출력되는 반면에 decoder은 확률값으로 나타나개 때문에 어떤 token이 나올지를 예측한다는 데에서 차별점을 지닌다. 전이 중에 구조화된 텍스트 입력을 단일 연속 토큰 시퀀스로 처리하는 순회 접근 방식에서 파생된 작업별 입력 적응을 활용한다. 이러한 적응은 사전 훈련 모델의 아키텍처를 최소한으로 변경하며 효과적으로 세밀 조정할 수 있도록 한다. 해당 논문에서, 준지도 학습 방식을 통해 언어 이해 과업들을 사용한다. 비지도 방식의 pre-training과 지도방식의 fine-tuning의 결합을 통해 훈련하는 것이다. 두 단계의 훈련 절차로 나눠볼 수 있다. 먼저, 미분류 데이터에서 언어 모델링을 사용함으로써 신경망 모델의 초기 매개 변수를 학습한다. 그런 다음 지도 학습을 이 파라미터들을 상응하는 지도 객체를 활용하여 타겟 태스크에 적용한다.

텍스트, 스크린샷, 번호, 라인이(가) 표시된 사진

자동 생성된 설명

**모델의 주로 사용 가능한 태스크는 무엇인가?**

다양한 자연어 이해 작업에 사용할 수 있다. NLI에 활용되어 모순되거나 대등한 문장의 관계를 인식하고 분류할 수 있다. 질의 응답과 상식 추론에 활용될 수 있다. 또 두 문장이 의미적으로 동일한지를 추론하는 텍스트 유사도 평가에도 활용될 수 있다. 마지막으로 텍스트를 분류하는 작업에도 활용될 수 있다.

**모델의 의의와 한계점이 무엇인가?**

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

해당 모델은 기존의 rnn과 같은 구조를 탈피하여 transformer 구조를 사용했다는 것에 큰 의의가 있다. transferring embedding들이 layer마다 최대 9%의 성능 향상을 일으킨다는 것을 밝혔다. 이는 pre-trained model에서 각각의 레이어들이 유용한 함수를 포함하고 있음을 나타낸다. 또한 transformer의 attentional memory가 기존 LSTM들에 비해 전이 시에 도움이 되며, 학습 횟수에 따라 꾸준히 성능이 증가하는 것을 보였다.

그러나 비지도 학습을 지향했음에도 fine-tuning과정과 input transformation이 들어갔다는 점에서 supervised가 사용된다는 한계가 있다.