



30913 유다희 - vscode 웹 제작

30918 이채은, 30927 황나영 - ppt 제작

GPT 모델



GPT 모델: transformer의 **decoder블록 만을 사용**한 자연어처리 모델로, 텍스트 데이터의 패턴을 학습하고 새로운 텍스트를 생성 하는데 사용한다.

GPT1 → GPT2 → GPT3로 발전해 오면서 엄청난 성능을 내기 시작

G: Generative

P: Pretrained

T: Transformer

-> 트랜스포머 모델에 기반한 사전훈련된 생성 모델

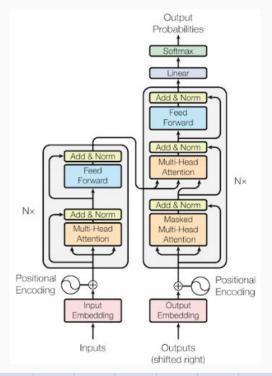
▼ > Chat GPT는 GPT-3.5모델을 강화학습을 통해 파인튜닝한 모델이다.

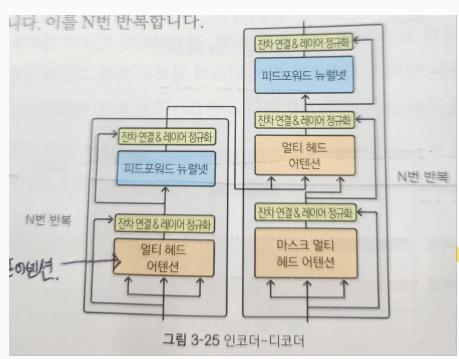
GPT 3.5 모델 정리



- 1. GPT 모델은 트랜스포머 모델에 기반한 사전훈련 생성 모델임.
- 2. 최대 2048개 토큰까지 입출력됨.
- 3. 12288차원으로 임베딩된 시퀀스를 입력 받음.
- 4. GPT3.5부터는 지도학습을 시작하였으며, 기존의 10배 이상의 데이터셋 학습, 파인튜닝하여 스스로 학습하게 함.

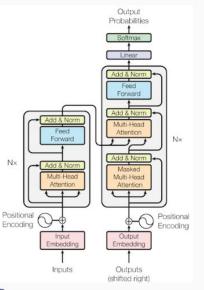








트랜스포머 모델: 자연어 처리 작업에서 우수한 성능을 보이는 모델로 **'셀프 어텐션'**이라는 핵심 메커니즘을 사용한다.



트랜스포머 모델은 왼쪽 사진에서 2개의 큰 블록으로 구성되어있다.

encoder - 왼쪽 블록 decoder - 오른쪽 블록

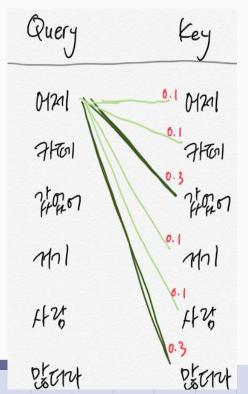
encoder의 역할: 입력 시퀀스(순서가 있는 문장)를 의미적인 표현으로 변환하는 역할수행

decoder의 역할: 인코더에서 생성된 의미적인 표현을 기반으로 출력 시퀀스를 **'생성'**하는 역할 수행

인코더는 입력 시퀀스를 특성 벡터로 변환하여 의미를 인코딩하고, 디코더는 인코더의 출력과 셀프 어텐션을 활용하여 출력 시퀀스를 생성한다.







블록의 구성요소

각 블록의 입력은 토큰화 → 임베딩 + 위치정보 의 처리 과정을 거친다.

BPE토큰화 - 많이 등장하는 문자열을 병합해 사전 형식으로 데이터 압축.

임베딩 - 토큰화된 자연어를 기계가 이해할 수 있는 벡터로 변경. 이 벡터들은 어텐션에서 각 시퀀스간의 유사도를 측정하는데 사용.

어텐션층 : 입력 시퀀스의 각 단어가 다른 모든 단어와 얼마나 관련이 있는지를 계산하는 메커니즘이다.

이것은 모델이 문장 내에서 문맥을 이해하고, 단어 간의 관계를 파악하는 데 도움이 된다.



Chat GPT의 언어생성 과정

1. 문장입력

2. 문장 **토**큰화

GPT-3.5모델은 50257개의 토큰이 등록되어 있기때문에 한단어가 (1,50257)의 행렬로 표현

3. 임베딩 + 위치정보

GPT-3.5모델은 12.288차원의 벡터를 입력으로 받음

4. 셀프 어텐션층

각 단어별로 중요도에 따른 가중치를 얻고 각 단어의 밸류와 가중합을 통해 출력 벡터 생성

5. 잔차연결, 정규화

출력+입력 으로 데이터의 정보를 보존하고, 정규화를 통해 출력의 분포 안정화

6. 인공신경망

각 시퀀의 벡터가 독립적으로 인공신경망을 통과 비선형 변환을 수행하여 모델이 더 복잡한 패턴학습

7. 잔차연결, 정규화2

출력+입력 으로 데이터의 정보를 보존하고, 정규화를 통해 출력의 분포 안정화

4~7의 과정이 96번 반복

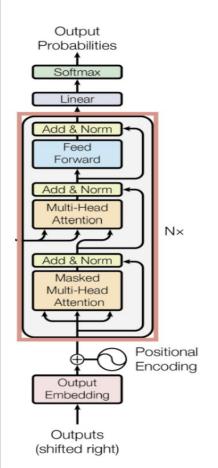
8. linear층

비선형 데이터를 선형데이터로 변환 > 각 단어의 임베딩을 어휘사전 크기의 벡터로 변환

9. softmax층

softmax층을 통과하면 50257개의 각 단어가 다음 단어로 선택될 확률 분포를 생성 가장 확률이 높은 단어 출력

출력된 단어는 전의 입력문장 맨뒤에 포함 되어 다시 입력으로 들어감











6.향후 개선점, 기대점 (활용법, 기대효과)







7. 마무리 인사

(조원들 느낀점)



