

DialoGPT

이름 / 학번 : 황민규 / V2022117

전공 : 메타버스 테크놀로지

DialoGPT는 2005년부터 2017년 기간동안 Reddit comment chain에 추출한 1억4천7백만 대화형 코멘트에 학습되어 automatic과(PPL같은) 싱글턴 대화 환경에서 human evaluation을 사람과 비슷한 성능을 얻는 것을 목표로 한 언어모델이다.

GPT-2

GPT-2는 transformer의 decoder 블록으로 구성된다. decoder block에 입력되는 입력벡터는 각 디코더셀의 self-attention 과정을 거친 뒤, 신경망 레이어를 지난다. 모든 디코더 블록을 거친 최종 결과물은 입력값에 대한 최종 셀프 어텐션 값을 가지고 있다. 이를 우리가 가진 임베딩 벡터와 곱해주면, 각 단어가 다음 단어로 등장할 확률값이 나오게 된다. 이 중에서 가장 확률값이 높은 것이 출력값이 되며, 또 다음의 입력값이 된다.

1. Introduction

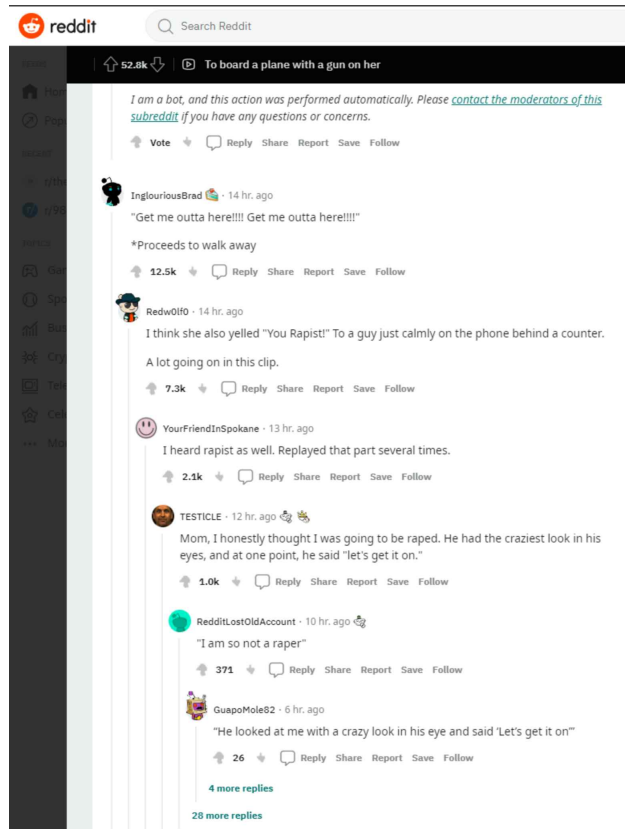
DIALOGPT는 GPT-2를 확장하여 conversational neural response generation을 완수하는것을 목표로 한다. neural response generation란 text-generation의 하위분야로서 자연스럽게 보이는 텍스트를 생성한다는 목표를 가지고있다.

DIALOGPT는 GPT-2와 같은 autoregressive(AR) language model이다. 모델 구성으로는 multi-layer transformer를 사용한다. 그러나 GPT-2와의 차이점은 Reddit discussion chain에서 추출된 대규모 대화 pair/session에서 학습한다는 점이다. 이러한 차이점의 대한 이유는 대규모 대화 pairs/session 학습을 통해 DIALOGPT가 대화 플로우에서 $P(\text{Target}, \text{Source})$ 에대한 joint distribution를 캡처할 수 있게 하는데에 있다.

2. Dataset

데이터셋은 2005년부터 2017년 기간동안 Reddit comment chain에 추출한 1억4천7백만 대화형 코멘트이다.

Reddit discussion은 스레드에 대한 응답 스레드는 하위 스레드의 루트 노드를 형성하기 때문에 자연스럽게 tree-structured한 응답 체인으로 펼쳐질 수 있습니다. 논문에서는 root node에서 leaf node까지의 각각의 path(하위 스레드)를 대화의 멀티턴을 가진 학습 인스턴스로 사용한다.



reddit 예시

데이터를 필터링 하는 기준은 아래와 같다.

- (1) URL이 있는 source나 target
- (2) 3개 이상의 단어 반복이 target에 포함된 경우
- (3) 응답이 가장 자주 사용하는 top 50 단어에 적어도 하나 이상 포함하지 않은 경우(예를 들면 the, of, a)(이는 영어가 아닐 수도 있기 때문)
- (4) 응답에 "[" 또는 "]"이 포함 된 경우 (이는 markup 언어 일 수도 있기 때문)
- (5) source와 target 시퀀스가 합쳐서 200 단어보다 긴 경우
- (6) target이 offensive language를 포함한 경우 (대용량 blcoklist에 매칭하는 방법으로 필터링)
- (7) 하위 레딧에 많은 수가 offensive한 내용을 포함할 가능성이 많다고 인식되는 경우
- (8) 단조로운 문장 적극적으로 배제 (1,000번 이상 본 tri-gram의 90%가 포함된 응답)

필터링 후 데이터 세트는 총 18억 개의 단어로 147,116,725개의 대화 인스턴스로 구성되었다.

3. Method

3.1 Model architecture

우리는 첫 번째로 대화세션 안에서 모든 대화 turns을 concat 시켜 긴 $\text{text}_{x_1, \dots, x_N}$ (N은 시퀀스 길이)를 만들고 끝에는 end-of-text-token을 넣는다.

source sentence는 $S = x_1, \dots, x_m$ 으로 target sentence (ground truth response)는 $T = x_{m+1}, \dots, x_N$ 으로 표기한다.

이때, 조건부 확률 $P(T|S)$ 은 조건부 확률의 일련의 곱으로 아래의 식과 같이 쓰여진다.

$$p(T|S) = \prod_{n=m+1}^N p(x_n | x_1, \dots, x_{n-1})$$

DialoGPT는 GPT-2와 동일하게 multi-turn dialogue를 하나의 text로 간주한다. 따라서 multi-turn dialogue session인 T_1, \dots, T_k 은 $p(T_k, \dots, T_2 | T_1)$ 로 볼 수 있고 이는 사실 $p(T_i | T_1, \dots, T_{i-1})$ (여기서 i 는 $m+1$) 조건부 확률을 product한 것이다. 결과적으로 $p(T_k, \dots, T_2 | T_1)$ 을 최적화하는 것은 모든 $p(T_i | T_1, \dots, T_{i-1})$ source-target 페어를 최적화하는 것이다.

3.2 Mutual Information Maximization

$$\hat{T} = \arg \max_T \{ \log p(T|S) \}$$

위의 공식처럼 일반적인 likelihood는 'I don't know, i'm ok'와 같이 어정쩡한 응답을 생성하게 한다고 추측한다. 이 공식은 target에서 source가 아닌 source에서 target에 대해서만 선택하기 때문이다. 오픈 도메인 텍스트 생성 모델의 bland하고 uninformative 함을 해결하기 위해, objective function을 maximum mutual information(MMI) scoring function으로 대체한다. MMI에서 파라미터는 아래의 공식처럼 source S 와 target T 간에 상호 정보(mutual information)를 최대화하도록 학습된다.

$$\hat{T} = \arg \max_T \{ (1 - \lambda) \log p(T|S) + \lambda \log p(S|T) \}$$

MMI는 pre-trained backward model(pre-trained language model)을 활용하여 주어진 응답에 대해 source sentence를 예측한다(ex. $P(\text{Source}|\text{target})$). 먼저 top-K 샘플링을 사용하여 hypotheses 셋을 생성한다($P(\text{Hypothesis}|\text{Source})$). 그리고 모든 hypotheses를 re-rank 하기 위해 $P(\text{Source}|\text{Hypothesis})$ 조건부 확률을 사용한다. 직관적으로 backward model(pre-trained model) likelihood를 최대화하는 것은 bland hypotheses에 불이익을 준다. bland 한 hypotheses는 많은 source query이 가능하기 때문에 모든 쿼리에 대해 확률값이 낮아지기 때문이다. 따라서 특정 쿼리에 대해 높은 확률 값을 가지지 않게 된다.