DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation

Created	@Feb 09, 2020 11:51 AM
Domain	https://arxiv.org/abs/1911.00536
Materials	https://arxiv.org/abs/1911.00536
Tags	deeplearning nlp
github	https://github.com/microsoft/DialoGPT

DIALOGPT(Dialogue generative pre-trained transformer)란?

- 크고 조정 가능한 신경 대화형 응답 생성 모델
- 2005~2017년 Reddit 코멘트 체인에서 추출된 1억 2천 5백만개의 대화식 교환에 대해 훈련된 Hugging Face PyTorch Transformer를 확장
- Single-turn dialogue settings에서 자동 및 인간 평가 측면에서 인간과 가까운 성능을 달성(?)
- 사전 훈련된 모델 및 교육 파이프라인이 공개되어 신경 반응 생성 및 보다 지능적인 오픈 도메인 대화 시스템 개발에 관한 연구가 쉬움

소개

- Reddit 데이터에 대해 학습된 대화형 응답 생성을 위한 조정 가능한 기가워드 스케일 신경망 모델
- 최근에 Transformer 기반 아키텍쳐 (Radford et al., 2018; Devlin et al., 2019; Raffel et al., 2019)를 사용한 대규모 사전 훈련의 발전이 성공적으로 입증

    - GPT-2가 매우 큰 데이터 세트에 대해 유창하고 어휘적으로 다양하며 내용이 풍부한 텍스트를 생성할 수 있음을 보여주었음.
    - 미세한 세밀도로 텍스트 데이터를 캡처하고 인간이 쓴 실제 텍스트를 면밀하게 모방한 고해상도 출력을 낼수 있는 능력이 있음.
- DIALOGPT는 대화형 신경 반응 생성 문제(the challenges of conversational neural response generation)를 해결하기 위해 GPT-2를 확장함.
    - 신경 반응 생성 : 프롬프트와 관련이 있는 자연적으로 보이는 텍스트 생성의 목적을 공유하는 텍스트 생성의 하위 범주
- 신경기계번역과 같은 텍스트 생성 과제보다 텍스트 요약 및 편집과 같은 일대다 문제를 풀어보자 라는 게 목표
- 대부분 open-domain neural response generation system은 내용 또는 스타일의 불일치나 장기적 문맥 정보의 부족 등의 문제가 있음
    - 이러한 것은 transformer 모델을 사용하면 장기 의존성 정보를 시간이 지남에 따라 더 잘 보전할 수 있으며, 깊은 구조보다는 대규모 데이터 세트를 활용하는게 더 효과적
- 그래서 DIALOGPT는 Auto regresive 언어 모델로 구성되며 multi-layer transfomer 아키텍처.
- GPT-2와 다르게 Reddit 토론 체인에서 추출한 대규모 대화 쌍/세션에 대해 train 함
    - 따라서 DIALOGPT가 더 세분화 된 대화 흐름에서 P(target;source)의 공동 분포를 포착 할 것이라고 가정함.
- 공개 벤치마크 데이터셋(DSTC-7)과 Reddit 게시물에서 추출한 새로운 6k 다중 참조 테스트 데이터 세트로 사전 훈련된 모델을 평가하여 SOTA를 찍었음.
- DIALOGPT는 새로운 데이터 세트(훈련 사례가 거의 없는 데이터 세트)에서 쉽게 활용 및 적용 할 수 있음.

Dataset

- 2005 ~ 2017년까지 Reddit에서 스크랩 된 comment 체인
- 필터링 조건
    - 소스 또는 URL
    - 3개 이상 반복 단어 포함

    - 응답에 1개 이상의 단어가 포함되지 않은 경우
    - 50개의 가장 빈번한 영어 단어가 적어도 하나 이상 포함 되어 있지 않은 경우
    - [,] 와 같은 특수 마커가 있는 경우
    - 소스 및 대상 시퀀스가 200단어 보다 길 경우
    - 블랙 리스트에 들어가 있는 공격적인 언어가 포함된 경우
- 필터링 후 18억 단어로 147,116,725(1억이상)개의 대화 인스턴스로 구성

Method

Model Architecture

- GPT-2 아키텍처를 기반
- GPT-2에서 12-24 layer transfomer, 직접 수정한 모델 깊이의 초기화 방식, 토큰나이저용 바이트 페어 인코딩을 상속
- multi-turndialogue session을 긴 테스트로 모델링하고 생성 작업를 언어 모델링으로 구성
    - 대화 세션 내의 모든 dialogue들을 긴 텍스트 x1~ xN(N은 시퀀스 길이)로 연결
    - Source 문장은 S = x1~xm
    - Target 문장은 T = xm+1~xN

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

    - multi-turn dialogue instances는 T1~TK 즉, p(TK,...,T2|T1)으로 쓸 수 있음
        - 이는 p(Ti|T1,...,Ti-1)읜 조건부 확률의 곱
    - single objective인 p(TK,...,T2|T1)를 최적화 하는 것은 p(Ti|T1,...,Ti-1) source-target 쌍을 최적화 하는 것으로 생각할 수 있음.

Mutual InformationMaximization

- Open-domain 텍스트 생성 모델은 단순하고 비정보적인 샘플을 생성하는 것으로 악명이 높음.
    - 이 문제를 해결하기 위해 Maximun Mutual Information(MMI) 스코어링 기능을 구현
    - MMI는 사전 훈련된 backward 모델을 사용하여 주어진 응답, p(Source|Target)에서 소스 문장을 예측합니다.
        - 먼저 top-K 샘플링을 사용하여 일련의 가설을 생성
        - P(Source|Hypothesis)의 확률을 사용하여 모든 가설을 재조정. (Hypothesis: 위에서 만든 가설)
        - 빈번하고 반복적인 가설이 많은 쿼리와 연관 될 수 있기 때문에 Maximizing backward model likelihood이 저런 가설들에게 불이익을 주어 특정한 쿼리에 대한 확률을 낮춰줌
        - Zhang et al에 따르면 샘플 평균 기준선으로 policy gradient를 사용하여 보상 R를 정의하면 ⇒ P(Source|Hypothesis)를 최적화하려고 시도
            - 검증된 보상은 안정적으로 개선 될 수 있지만 RNN 아키텍처에 따른 훈련과 달리 강화 학습(RL) 훈련은 locally-optimal solution으로 쉽게 수렴 되는 것으로 나타났는데, 여기서 가설은 단순히 Source 문장을 반복하고 상호 정보가 극대화 됨
            - 따라서 transformer가 강력한 모델 표현력때문에 local optima에 쉽게 갇힐 수 있다고 가정함
                - 정규 RL 훈련에 대한 조사는 향후 작업으로 넘김.

Result

Experimental Details

- 총 117M, 345M, 762M의 총 매개 변수를 사용하여 3가지 크기의 모델을 학습. 모델 사양은 아래와 같음

Model	Layers	D_{emb}	B
117M	12	768	128
345M	24	1024	64
762M	36	1280	32

  - B는 GPU당 배치 크기
- 이 모델은 50,257개의 어휘를 사용하며 NVLINK를 사용하여 16개의 Nvidia V100으로 훈련
  - 16,000 warm-up step으로 Noam 학습 rate 스케줄러를 사용
  - 학습 속도는 validation loss를 기준으로 선택
  - 각 모델은 validation loss가 진행되지 않을때 까지 학습
    - 중소형 모델은 최대 5 epochs
    - 대형 모델은 최대 3 epochs

Speeding up training

- 훈련 과정을 가속화하고 GPU 메모리 제한을 수용하기 위해 모든 훈련 데이터를 lazy-loading database file로 압축하여 필요할 때만 데이터가 로드되는 식으로 작업
- 훈련을 확장하기 위해 별도의 비동기 데이터 프로세스를 확용
  - 훈련 시간은 대략적으로 GPU의 수와 선형적으로 감소
  - 유사한 길이의 대화를 동일한 배치로 그룹화 하여 동적 배치 전략을 채택

DSTC-7 Dialogue Generation Challenge

- The DSTC (Dialog System Technology Challenges) 7 track (Galley et al., 2019)의 목표는 외부 지식에 근거한 정보를 주입하여 채팅을 넘어서는 대화 응답을 생성하는 것이 목표인 end-to-end 대화 모델링
  - 이 task는 구체적이거나 사전 정의된 목표가 없다는 점에서 목표 지향적, task 지향적, task 완료 대화로 생각하는 것과는 다르다

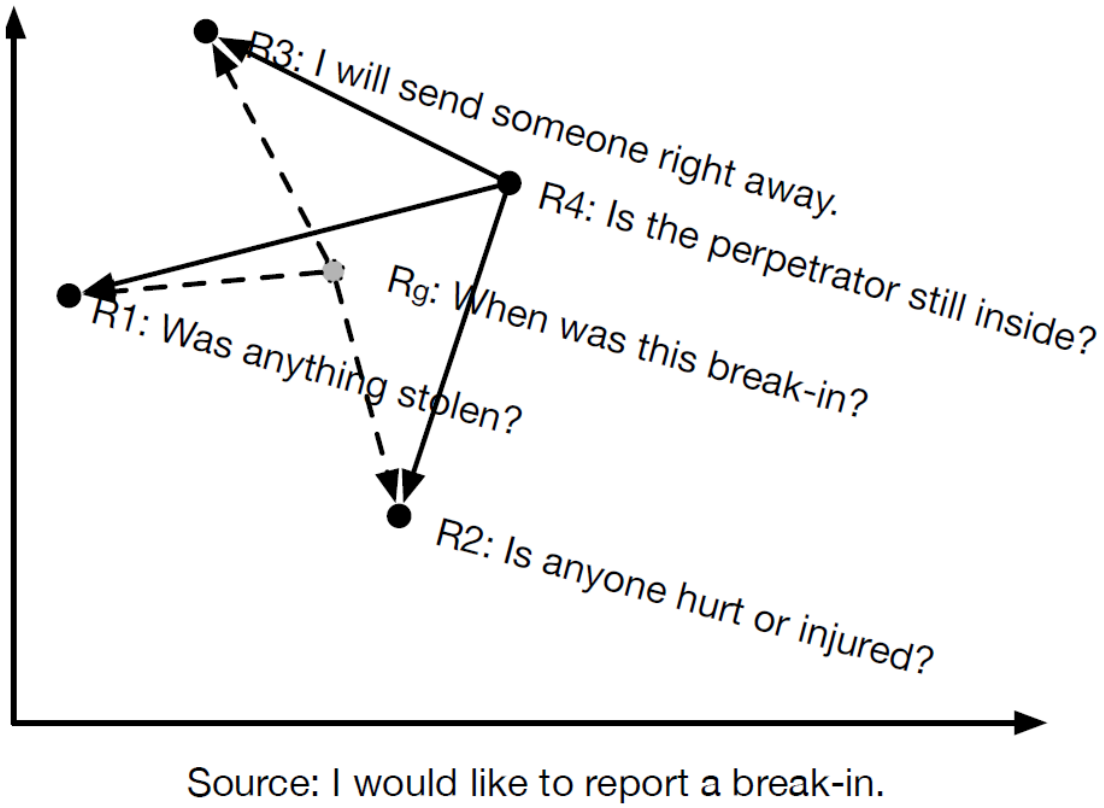
    - 대신에 사람들이 정보를 공유하는 직장, 그 밖의 생산적인 환경에서 보이는 것과 같이 기본적인 목표가 사전에 잘못 정의되거나 알려지지 않은 인간과 같은 상호 작용이 목표
- 테스트 데이터는 Reddit 데이터의 대화 스레드를 포함
    - Multi-reference test set를 만들기 위해 6개 이상의 응답이 포함된 대화 세션을 활용
    - turn 길이와 같은 다른 필터링 기준을 고려할 때 2208 사이즈의 5-reference test set을 생성. (각각의 경우, 이 task에서 사람의 성과를 평가하기 위해 6명의 사람 응답 중 하나가 따로 설정 되어 있음). 훈련 데이터는 테스트 세트와 다른 시간의 범위에서 수집
- BLEU, METEOR, NIST를 포함한 표준 기계 번역 메트릭을 사용하여 자동 평가를 수행
    - NIST는 정보 이득에 의해 n-그램과 가중치가 일치하는 BLEU의 변경. 즉, 정보가 없는 n-그램에 간접적으로 불이익을 줌
- 어휘 다양성을 평가하기 위해 엔트로피와 Dist-n을 사용함.
- DIALOGPT를 두가지 기준으로 비교
    - 사내 경쟁 시퀀스-to-시퀀스 모델인 PERSONALITYCHAT
    - 마이크로스프트 Azure의 인지 서비스로 생산에 이용되어 온 트위터 데이터에 대한 훈련 실시

Method	NIST N-2	NIST N-4	BLEU B-2	BLEU B-4	METEOR	Entropy E-4	Dist D-1	Dist D-2	Avg Len
PERSONALITYCHAT	0.19	0.20	10.44%	1.47%	5.42%	6.89	5.9%	16.4%	8.2
Team B	2.51	2.52	14.35%	1.83%	8.07%	9.03	10.9%	32.5%	15.1
Ours(117M)	1.58	1.60	10.36%	2.02%	7.17%	6.94	6.2%	18.94%	13.0
GPT(345M)	1.78	1.79	9.13%	1.06%	6.38%	9.72	11.9%	44.2%	14.7
Ours(345M)	2.80	2.82	14.16%	2.31%	8.51%	**10.08**	9.1%	39.7%	16.9
Ours(345M,Beam)	**2.92**	**2.97**	**19.18%**	**6.05%**	**9.29%**	9.57	**15.7%**	**51.0%**	14.2
Human	2.62	2.65	12.35%	3.13%	8.31%	10.45	16.7%	67.0%	18.8

Table 2: DSTC evaluation. "Team B" is the winner system of the DSTC-7 challenge. "Beam" denotes beam search. "Human" represents the held-out ground truth reference.

- 위 표에 345M 파라미터와 Beam search를 갖춘 DIALOGPT가 대부분 측정에서 높은 점수를 획득
- 345M 파라미터가 117M보다 전반적으로 좋다

- Beam Search (너비 10)은 BLEU, DIST 점수를 크게 개선하고 NIST, METEOR를 약간 개선
- Source-Target 쌍으로 미세조정 되며 DSTC 훈련 세트의 기초 정보를 활용하지 않음
    - 사전 훈련 과정에서 풍부한 배경 정보를 학습하고 기본 문서가 없기 때문에 방해가 되지 않음
- DIALOGPT의 점수는 사람의 점수보다 높음
    - 인간보다 현실적이라기 보다는 대화의 외톨이적 성격 탓



- 다수의 사람 반응 (R1-R4)은 Source 표현에 잘 일치 함
- 일반성을 상실하지 않는 한, R1-R3은 테스트할 기본적 진실을 참조인 반면, R4는 사람 점수를 계산하는 역할을 하는 지속적 인간 응답이라고 가정

- 의미 공간에서 잘 훈련된 모델에서 생성된 응답 Rg는 아마도 모든 가능한 반응의 기하학적 중심 근처에 놓여 있는 경향이 있음
    - 훈련의 목표는 가장 가능성이 높은 응답을 생성하기 때문

- 모든 훈련 인스턴스의 기하 평균에 가깝기 때문에 이러한 인스턴스를 평균화
- 생성된 Rg는 표준화 된 인간 응답 R4보다 R1-R3으로부터 더 낮은 의도적 거리(BLEU와 같은 더 높은 점수에서 관리됨)을 가짐

A New Reddit Multi-reference Dataset

Method	NIST N-2	NIST N-4	BLEU B-2	BLEU B-4	METEOR	Entropy E-4	Dist D-1	Dist D-2	Avg Len
PERSONALITYCHAT	0.78	0.79	11.22%	1.95%	6.93%	8.37	5.8%	18.8%	8.12
Training from scratch:									
Ours(117M)	1.23	1.37	9.74%	1.77%	6.17%	7.11	5.3%	15.9%	9.41
Ours(345M)	2.51	3.08	16.92%	4.59%	9.34%	9.03	6.7%	25.6%	11.16
Ours(762M)	2.52	3.10	17.87%	5.19%	9.53%	9.32	7.5%	29.3%	10.72
Training from OpenAI GPT-2:									
Ours(117M)	2.39	2.41	10.54%	1.55%	7.53%	10.77	8.6%	39.9%	12.82
Ours(345M)	3.00	3.06	16.96%	4.56%	9.81%	9.12	6.8%	26.3%	12.19
Ours(345M, Beam)	**3.4**	**3.5**	**21.76%**	**7.92%**	10.74%	10.48	**12.38%**	**48.74%**	11.34
Ours(762M)	2.84	2.90	18.66%	5.25%	9.66%	9.72	7.76%	29.93%	11.19
Ours(762M, Beam)	2.90	2.98	21.08%	7.57%	10.11%	10.06	11.62%	44.07%	10.68
Ours(345M, MMI)	3.28	3.33	15.68%	3.94%	**11.23%**	**11.25**	9.39%	45.55%	17.21
Human	3.41	4.25	17.90%	7.48%	10.64%	10.99	14.5%	63.0%	13.10

Table 3: 6K Reddit multi-reference evaluation. "Beam" denotes beam search. "Human" represents the held-out ground truth reference.

- 5k 예제를 가진 multi-reference test set 결과
- GPT-2를 사전 훈련된 모델로 사용하여 처음부터 훈련 및 미세 조정의 두가지 설정하여 테스트
  - 두 설정 모두 큰 모델이 작은 모델보다 성능이 우수
  - 사전 훈련된 GPT-2 모델로 fine-tuning 된 것과 scratch로 훈련된 것을 비교하면 작은 모델에서는 GPT-2를 사용한 것이 성능에 도움이 됨
  - 사람 점수보다 345M, Beam이 BLEU 점수가 더 높게 나왔음.

Re-ranking The Response Using MMI

- GPT-2 중간 모델에서 fine-tune된 345M 모델을 사용하여 top-K(K=10)을 사용하여 각 입력 소스 문장에 대해 16개 샘플을 생성
- GPT-2 중간 모델에서 fine-tune된 345M 모델인 backward 모델을 사용하여 다시 순위를 매기는 단계가 이어짐
- 역행 모델 손실이 가장 적은 반응이 평가를 위해 선택되어짐 (표3 MMI)
- Greedy 생성과 비교할 때 MMI re-ranking화는 높은 NIST, METEOR, Entropy 및 Dist 점수로 더 다양한 응답을 생성하지만 BLEU에서는 약간 감소

Generation Examples

source	Response
who is the first president of the United States?	George Washington
what is the boiling point of water?	I think it's about 212 F.
which one is bigger, sun or moon?	The sun .
which animal has black and white stripes?	A black and white striped cat .

Table 4: Addressing commonsense questions

Role	response
User	Does money buy happiness?
Bot	Depends how much money you spend on it .
User	What is the best way to buy happiness ?
Bot	You just have to be a millionaire by your early 20s, then you can be happy .
User	This is so difficult !
Bot	You have no idea how hard it is to be a millionaire and happy . There is a reason the rich have a lot of money

Table 5: An interactive example of multi-turn dialogue

Role	Response
User	what is the meaning of life ?
Bot	The meaning is to be with your family and friends .
Bot	I'm going to guess : It means that your parents and friends have loved you . That's the meaning , I believe .
Bot	I think the meaning is to make sure you're alive when you die . So you'll be with your family , friends , and yourself .
Bot	It's not , but thanks for trying to help .

Table 6: An example of multi-turn self-playing dialogue with user prompt

- 표 4 (대화식 채팅) 및 표 5 (사용자 프롬프트가 포함 된 자체 재생 봇)에 샘플 생성 대화가 제공
- 출력은 top-K 샘플링을 기반으로 하며 10개의 랜덤 샘플에서 수동으로 응답을 선택
  - DIALOGPT는 Reddit 데이터에서 배울 수 있는 풍부한 정보로 인해 상식 문제를 어느정도 해결할 수 있는 능력을 보여줌
  - 경우에 따라 원하는 답변 대신에 대안적이고 합리적인 답변을 생성
  - RNN보다 멀티 턴 생성을 더 잘 처리 할 수 있으며 관련하여 보다 일관된 경향이 있음(표5)

Human Evaluation

- 크라우드 소싱을 사용하여 Reddit 6K 테스트 데이터 세트에서 무작위로 샘플링된 2000개의 테스트 소스를 평가
  - 시스템을 쌍으로 구성하여 각 시스텐 출력 쌍을 무작위로 3명의 심사원에게 제시
  - 심사원은 목적 적합성, 정보성 및 생성물이 3점 Likert와 같은 척도를 사용하여 얼마나 인간적인지 평가
  - 심사위원은 자격 테스트를 통과해야하고 스팸 탐지 제도가 시행.

Relevance: *A and B, which is more relevant and appropriate to the immediately preceding turn?*				
System A		Neutral	System B	
DialoGPT (345M)	3281 (**72%**)	394 (9%)	882 (19%)	PersonalityChat ****
DialoGPT (345M)	2379 (40%)	527 (9%)	3094 (**52**%)	DialoGPT (345M, w/ MMI) ****
DialoGPT (345M)	3019 (**50%**)	581 (10%)	2400 (40%)	DialoGPT (345M, Beam) ****
DialoGPT (345M)	2726 (**45%**)	576 (10%)	2698 (45%)	DialoGPT (762M)
DialoGPT (345M)	2671 (45%)	513 (9%)	2816 (**47**%)	Human response
DialoGPT (345M, w/ MMI)	2871 (**48%**)	522 (9%)	2607 (43%)	Human response ***

Informative: *A and B, which is more contentful, interesting and informative?*				
System A		Neutral	System B	
DialoGPT (345M)	3490 (**77%**)	206 (5%)	861 (19%)	PersonalityChat ****
DialoGPT (345M)	2474 (41%)	257 (4%)	3269(**54**%)	DialoGPT (345M, w/ MMI) ****
DialoGPT (345M)	3230 (**54%**)	362 (6%)	2408(40%)	DialoGPT (345M, Beam) *****
DialoGPT (345M)	2856 (**48%**)	303 (5%)	2841(47%)	DialoGPT (762M)
DialoGPT (345M)	2722 (45%)	234 (4%)	3044(**51**%)	Human response ****
DialoGPT (345M, w/ MMI)	3011 (**50%**)	234 (4%)	2755(46%)	Human response **

Human-like: *A and B, which is more likely to be generated by human rather than a chatbot?*				
System A		Neutral	System B	
DialoGPT (345M)	3462 (**76**)%	196 (4%)	899 (20%)	PersonalityChat ****
DialoGPT (345M)	2478 (41)%	289 (5%)	3233 (**54%**)	DialoGPT (345M, w/ MMI) ****
DialoGPT (345M)	3233 (**54**)%	340 (6%)	2427 (40%)	DialoGPT (345M, Beam) ****
DialoGPT (345M)	2847 (**47**)%	321 (5%)	2832 (47%)	DialoGPT (762M)
DialoGPT (345M)	2716 (45)%	263 (4%)	3021 (**50**%)	Human response ***
DialoGPT (345M, w/ MMI)	2978 (**50**)%	241 (4%)	2781 (46%)	Human response *

Table 7: Results of **Human Evaluation** for relevance, informativeness and human-response possibility, showing preferences (%) for our model (DialoGPT) vis-a-vis its variants and real human responses. Distributions are skewed towards DialoGPT with MMI, even when compared with human outputs. Numbers in bold indicate the most preferred systems. Differences in mean preferences are statistically significant where indicated (* $p \leq 0.01$, ** $p \leq 0.001$, *** $p \leq 0.0001$, **** $p \leq 0.00001$).

- 위 표는 목적 적합성, 정보성 및 인간에 대한 전체적인 심사원의 선호도를 원시 숫자와 전체 비율을 표시
- PersonalityChat 보단 DialoGPT에 대한 강한 선호도를 확인 할 수 있음
- 바닐라 DialoGPT 중간 모델이 이미 인간의 반응 품질에 가까울 수 있음을 시사
  - 가끔 심사원들이 인간의 반응보다 MMI 변형을 선호 하는것도 보임
  - 아마 심사원들이 인터넷 밈에 연관된 것들이 익숙히 않아서 생긴 문제로 판단
- 중요도 시험과 사용된 인적 평가 템플릿을 포함한 추가 세부 사항은 부록에 수록

Related work

- 대규모 사전 훈련된 transformer 모델을 가진 open-source
  - Huggingface Conv-AI transfoer learning repository
    - GPT-2 Transformer 언어 모델을 기초한 전달 학습으로 대화형 AI 시스템을 훈련하기 위한 코드가 포함되어 있어 ConvAI-2 dialogure competition

에서 SOTA를 달성

- DLGnet
  - 대화 데이터셋에 대해 훈련된 대형 transformer로 multi-turn 대화 생성에서 우수한 성능 발휘
- AllenNLP
  - 대규모 사전 훈련된 bi-LSTM 문장 표현 학습 프레임 워크 ELMo를 포함하여 많은 자연 언어 처리 작업을 위한 툴킷으로 개발
- Texar
  - 스타일 전송 및 제어 가능한 세대를 포함한 텍스트 생성에 초점
  - 스퀀스 모델링 도구와 함께 강화 학습 기능이 포함
- DeepPavlov
  - Task 중신의 대화에 중점을 둔 인기 있는 툴
  - 공개 레포지토리는 질문 답변과 감정 분류를 위한 몇 가지 데모와 사전 훈련된 모델을 포함
- Icecaps
  - 인격 또는 외부 지식에 바탕을 둔 기술, 멀티태스킹 훈련과 같은 기술을 갖춘 응답 생성 툴킷
- ConvAI2 challenge
  - 개인화된 대화에 초점을 맞춤
- ParlAI
  - 업무 중심의 대화 시스템 개발을 위한 또 다른 library
  - 크라우드소싱된 데이터로 훈련된 지식 기반 채팅봇을 위한 사전 훈련된 모델을 포함
- Text-to-Text Transformer
  - 다중 텍스트 모델링 작업을 통합하고 다양한 자연어 생성 및 이해 벤치 마크의 최신 결과를 달성

Limitations and risks

- DIALOGPT는 모델로만 출시되며 디코더 구현의 책임은 사용자에게 있음

- 훈련전 명백한 공격 데이터의 양을 최소화하려는 노력에도 불구하고 DIALOGPT는 공격을 유발할 수 있는 출력을 생성할 수 있는 잠재력이 있음
- 산출물은 데이터에 내재된 성별 및 기타 과거 편견을 반영 될 수 있음
  - 따라서, 생성된 응답은 비윤리적, 편향적 또는 불쾌한 제안과 합의를 표현하는 경향이 보일 수 있음
  - 이러한 문제는 대규모 자연 발생 데이터셋에서 훈련된 최신 end-to-end 대화 모델에서 알려진 문제
- DIALOGPT를 만든 동기는 연구자들이 이러한 문제를 조사하고 완화 전략을 개발할 수 있도록 하는 것

Concolusion

- 대규모 실제 Reddit 데이터셋에 대한 훈련을 받은 open-domain 사전 훈련 모델을 만듬
- 분산 훈련 파이파라인과 몇 시간 안에 적절한 크기의 맞춤형 데이터셋에 대한 대화 모델을 얻을 수 있도록 fine-tune할 수 있는 몇가지 사전 훈련 모델로 구성
- DIALOGPT는 완전히 개방되어 있고 배포가 용이하여, 사용자가 미리 훈련된 대화 시스템을 다양한 데이터 세트를 사용하여 부트스트랩 훈련으로 그리고 새로운 애플리케이션과 방법론으로 빌딩 블록으로 확장할 수 있음
- 향후, 독성 발생을 검출하고 제어하는 방법을 조사하고 강화학습을 확용하여 발생된 반응의 관련성을 더욱 개선하고 모델이 터무니 없는 반응을 발생시키지 않도록 할 것임.

https://arxiv.org/abs/1911.00536