

TransferTransfor 논문

해당 논문은 GPT2를 활용해서 챗봇을 구현하였음.

Transfer - Transfer learning을 언급

Transfor - Transformer를 언급 (Transformer - RNN의 Long term dependency와 병렬 처리가 불가능한 부분에 대한 해결책 제서)

GPT2 관련 Input data

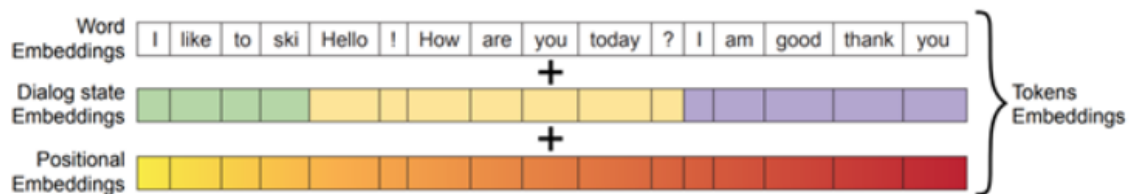


Figure 1: TransferTransfo's input representation. Each token embedding is the sum of a word embedding, a dialog state embedding and a positional embedding.

GPT, BERT는 3가지의 input embedding을 필요로 함

- Word Embedding (Token Embedding - W2V, Glove, BPE, Sentence Piece 등)
- Segment Embedding (위 예시에서 Dialog state Embedding)
 - Personality Sentence
 - Utterance from PERSON1
 - Utterance from PERSON2
- Positional Embeddings

※ Pretraining할 때에는 GPT2의 그것을 그대로 활용

※ Fine Tuning할 때에 우리의 dataset(논문에선 PERSONA-CHAT dataset)을 활용하고 augmented input representation(Dialog state Embedding을 활용하는 것 같음), 그리고 multi task learning scheme을 활용한다.

본 논문에서는 Segment Embedding을 3가지로 나눠서 이야기 한다. Personality, history(지금까지의 대화), Person2이 발화 **

Persona 1	Persona 2
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones

[PERSON 1:] Hi
[PERSON 2:] Hello ! How are you today ?
[PERSON 1:] I am good thank you , how are you.
[PERSON 2:] Great, thanks ! My children and I were just about to watch Game of Thrones.
[PERSON 1:] Nice ! How old are your children?
[PERSON 2:] I have four that range in age from 10 to 21. You?
[PERSON 1:] I do not have children at the moment.
[PERSON 2:] That just means you get to keep all the popcorn for yourself.
[PERSON 1:] And Cheetos at the moment!
[PERSON 2:] Good choice. Do you watch Game of Thrones?
[PERSON 1:] No, I do not have much time for TV.
[PERSON 2:] I usually spend my time painting: but, I love the show.

Table 1: Example dialog from the PERSONA-CHAT dataset. Person 1 is given their own persona (top left) at the beginning of the chat, but does not know the persona of Person 2, and vice-versa. They have to get to know each other during the conversation.

Model

Model은 기존 GPT2에다가 2개의 Linear 층을 병렬로 연결

(여기에서 기존 GPT2는 pretrained된 모델)

- 분류를 위한 Linear 층 (fake data vs real data)
 - Personality 에 부합하고, history와 어색하지 않게 Person2의 발화를 선정
- Language Model을 위한 Linear 층(for 단어 생성)
 - 문장이 비문법적으로 그리고 단어가 적절하지 않게 만들어지는 것을 방지

즉 Multi Task Learning임 - BERT와 비교하자면 분류를 위한 Linear 층은 Next Sentence Prediction에 부합하고, 두번째는 [MLM]에 부합한다고 할 수 있음.

Task Oriented Dialog System

1. intent classification task

ex) "I want to book a ticket from New York to San Francisco for tomorrow"

- intent : flight ticket booking

2. slot tagging

- extract relevant information(slots)

이 두 개를 CNN이든 RNN으로 활용한다.

- intent classification과 slot tagging을 다 Natural Language Understanding이라고 한다.

NLU를 통해서 얻은 정보는 Dialog Manager에서 활용한다.

이 때 Dialog Manager는 질문을 재차 물어봐야 되는 경우도 있다.

가령 "Book me a ticket from New York for tomorrow" 라고 하면

intent classification 에서 flight ticket booking을 파악한다. 그에 대응되는 slot은 3개임

(origin, date, destination) 이 때에 해당 발화는 2가지만 있기 때문에 origin에 대응되는 질문을 물어보아야 한다.