

연구학점제 발표 세미나

GPT2 기반 text generation 성능 관찰

2021.9.1~2021.12.10

배수영

Content

- Study overflow
- Concept
- Setup
- Experiments
- Results
- Conclusion

Study Overflow

<Study>

- GPT2, Text-generation study
- A Knowledge-Enhanced Pretraining Model for Commonsense Story Generation 논문 리뷰

<Project>

- ROCStories dataset 사용
- 1. GPT2 모델 학습 -> finding optimal hyperparameter using Perplexity
- 2. 문장을 생성 -> finding optimal hyperparameter using ROUGE, BLEU Score
- 어떤 문장이 생성되었는지, 원래 문장과 비교해보기 (반복성, 문법성, 일관성 확인)

Concept

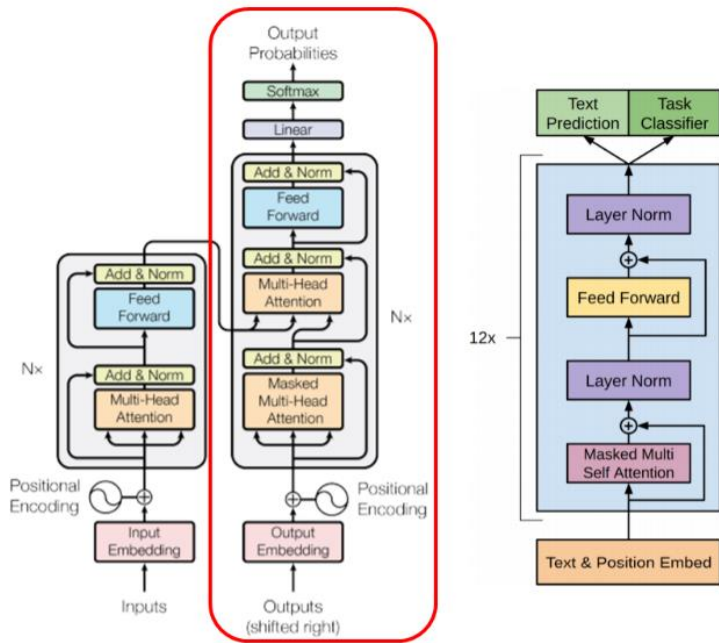
1. Text-generation

- 순환신경망(ex: RNN, LSTM)
- 입력, 출력 : 시퀀스 -> many-to-one 구조 사용
- 다음 단어의 확률을 예측할 때 auto-regressive 방법 사용한다.
- 다음 단어 선택하는 방법!
 1. Greedy Search : 가장 높은 확률 가지는 단어를 그 다음 단어로 사용
 2. Beam Search: 단어 선택을 beam 개수만큼 선택
 3. Sampling: 다음 단어에 대한 확률 분포 기반으로 선택
 4. TOP-K Sampling: 확률이 높은 k개 필터링하고 이를 기반으로 다시 확률 계산
 5. TOP-P Sampling: 누적 분포 계산, p 값 초과하는 단어들 sampling

Concept

2. GPT-2 (Language Models are Unsupervised Multitask Learners)

- 기존 GPT 모델과 동일한 모델 : transformer decoder 부분 + layer normalization 층 추가
- WebText(40GB) dataset 사용해서 학습
- Byte Pair Encoding(BPE) 형식의 입력을 사용.



Concept

3. Story Generation(A Knowledge-Enhanced Pretraining Model for Commonsense Story Generation)

- GPT2 사용해서 문장 생성 시 문제점: 단어 반복 등장, 관련 없는 내용, 문법성 파괴
- Solution: 외부 상식 지식(ConceptNet, ATOMIC) 적용, classification task 추가
- GPT2 모델 : batch size=10, learning rate=1e-4, Adam optimizer
- 이 논문에서 사용한 GPT2 모델 parameter 를 기본으로 하고 실험을 진행함.

Setup

1. Model

- GPT2LMHeadModel
- GPT2Tokenizer

```
#gpt-model setting
tokenizer = GPT2Tokenizer.from_pretrained('gpt2')
model = GPT2LMHeadModel.from_pretrained('gpt2')
```

2. Dataset

- ROCStories dataset
- ROCStories_winter2017 - ROCStories_winter2017.csv

Title		Five-sentence Story					
The Test		Jennifer has a big exam tomorrow. She got so stressed, she pulled an all-nighter. She went into class the next day, weary as can be. Her teacher stated that the test is postponed for next week. Jennifer felt bittersweet about it.					
The Hurricane		Morgan and her family lived in Florida. They heard a hurricane was coming. They decided to evacuate to a relative's house. They arrived and learned from the news that it was a terrible storm. They felt lucky they had evacuated when they did.					
Spaghetti		Tina made spaghetti for her boyfriend. It took a lot of work, but she was very proud. Her boyfriend ate the whole plate and said it was good. Tina tried it herself, and realized it					
	storyid	storytitle	sentence1	sentence2	sentence3	sentence4	sentence5
0	8bbe6d11-1e2e-413c-bf81-eaea05f4f1bd	David Drops the Weight	David noticed he had put on a lot of weight re...	He examined his habits to try and figure out t...	He realized he'd been eating too much fast foo...	He stopped going to burger places and started ...	After a few weeks, he started to feel much bet...
1	0beabab2-fb49-460e-a6e6-f35a202e3348	Frustration	Tom had a very short temper.	One day a guest made him very angry.	He punched a hole in the wall of his house.	Tom's guest became afraid and left quickly.	Tom sat on his couch filled with regret about ...
2	87da1a22-df0b-410c-b186-439700b70ba6	Marcus Buys Khakis	Marcus needed clothing for a business casual e...	All of his clothes were either too formal or t...	He decided to buy a pair of khakis.	The pair he bought fit him perfectly.	Marcus was happy to have the right clothes for...
3	2d16bcd6-692a-4fc0-8e7c-4a6f81d9efa9	Different Opinions	Bobby thought Bill should buy a trailer and ha...	Bill thought a truck would be better for what ...	Bobby pointed out two vehicles were much more ...	Bill was set in his ways with conventional thi...	He ended up buying the truck he wanted despite...

Setup

3. New_dataset

- (sentence1, sentence2)+(sentence2, sentence3)+(sentence3, sentence4)+ (sentence4, sentence5)
- Training 할 때 두 문장을 합쳐서 모델 input으로 설정.
- 총 210660 개 dataset 중 10000 random sampling

4. Train, Valid(=Test) split

- 9000개 train, 1000개 valid

Train, Valid dataset
sent
They begin to swim. They race each other.

Test dataset	
sentence1	Sentence2
They begin to swim.	They race each other.

Experiments

1. Train

1. Batch size= 4, 8, 16, 32
2. epoch=5, 10, 15, 20
3. Learning rate= $2e-5$, $5e-5$, $1e-4$, $5e-3$
4. Eps= $1e-8$
5. Loss, perplexity
->최적의 hyperparameter 찾기

Batch size	epoch	Learning rate	final Loss	perplexity
4	5	$1e-4$	0.1837	6.0758
8	5	$1e-4$	0.2755	8.839
16	5	$1e-4$	0.3151	12.456
32	5	$1e-4$	0.2530	11.375

Batch size	epoch	Learning rate	final Loss	perplexity
4	5	$1e-4$	0.1837	6.076
4	10	$1e-4$	0.1945	10.990
4	15	$1e-4$	0.3078	338.938
4	20	$1e-4$	0.2619	6.825

Batch size	epoch	Learning rate	final Loss	perplexity
4	5	$2e-5$	0.2124	11.539
4	5	$5e-5$	0.2390	10.907
4	5	$1e-4$	0.1837	6.0758
4	5	$5e-3$	0.2911	88.738

Experiments

2. Generate sentences

1. Top_p=0.85, 0.95, 0.99
2. Temperature=0.0, 0.2, 0.5, 1.0
3. Max length=300
4. 생성된 문장에서 \n, \t, ", <|endoftext|> 제외
5. Bleu, Rouge
-> 최적의 hyperparameter 찾기

Top_p	temperature	Blue-1	ROUGE-1 r	ROUGE-1 p
0.85	0	0.076	0.092	0.094
0.95	0	0.080	0.066	0.077
0.99	0	0.071	0.078	0.074

Top_p	temperature	Blue-1	ROUGE-1 r	ROUGE-1 p
0.85	0	0.076	0.092	0.094
0.85	0.2	0.078	0.082	0.106
0.85	0.5	0.076	0.091	0.113
0.85	1.0	0.078	0.079	0.085

Results

1. Train

Batch size	epoch	Learning rate	final Loss	perplexity
4	5	1e-4	0.1837	6.076

2. Generate

Top_p	temperature	Blue-1	ROUGE-1 r	ROUGE-1 p
0.85	0.5	0.076	0.091	0.113

Results

3. Generated Text (74개)

	Input sentence	Output generated sentence	repetition	logic	consistency
1	The pastor droned on with the service.	The pastor was not allowed to speak to the congregation.	O	O	O
2	When they all arrived and were about to start the police showed up.	They were told to leave.	O	O	O
3	Javier loved playing outside.	But he didn't like the heat.	O	O	X
4	My brothers and I fought, and one of us broke the vase.	He was a bad boy.	O	O	O
5	They decided to have dessert at a bakery.	They decided to have dinner with friends.	O	O	X
6	Jim was really into books.	She was a good teacher.	O	O	X
7	He would visit often.	His wife was a doctor.	O	O	X
8	Cathy wanted to plan a surprise party for Danny's special day.	He didn't know what to do.	O	O	X
9	They needed to get a family vehicle.	The driver of the car was also injured.	O	O	X
10	At first he was shy and scared of the baby.	Then he was scared of the baby.	X	O	O
11	Will decided to go to the beach anyway.	He was very disappointed.	O	O	X
12	The laptop began to beep and shut down.	He told the police he was going to be arrested.	O	O	X
13	Paul forgot to get his girlfriend a card for Valentine's Day.	0 During the walk, he dropped down on one knee.	O	X	O
14	The church had plenty of insurance coverage.	But the problem was that they were not insured.	O	O	O
15	Everybody was dancing but him.	He was like a bad biker.	O	X	O
16	He specifically hated squids though.	He was afraid they would be hurt.	O	O	O
17	Lacey and her boyfriend went to the fair.	They were told to wait for the police.	O	O	O
18	The woman picked up a small puppy from the liter.	She was very excited to see it.	O	O	O
19	I told my wife Google Maps has a smartphone app.	She loved it.	O	O	O
20	He got all of the ingredients.	He was so excited to get it.	O	O	O

Conclusion

- 의도된 문장과 생성된 문장이 달라 비교하기 어려운 점
 - > 원하는 문장이 생성되게 하기 위한 방법?
- 생성 모델에서 BLEU score과 Rouge 값이 낮게 나온다
 - > 번역 task에 주로 사용되는 지표로 비교하는 데 한계가 있다.
- Generation task 에 맞는 평가 지표가 추가적으로 필요할 것 같다.
 - > 문장 자체의 문법성, 반복성, 일관성 확인할 수 있는 평가지표?