

요약문을 또 생성해야 한다니
럭키비키잖아.. 🦴💖

Team I NLP 1팀

197기 심승현 이소희 이승준

CONTENTS

01

공모전 소개

- 목적
- 데이터
- 평가지표

02

모델 선정 과정

- KoBart v2

03

파인튜닝 과정

- 전처리
- 학습
- 생성

04

결과 기술

- 결과
- 한계 및 의의





01. 공모전 소개

01. 공모전 소개

대회 목적

‘일상 대화 요약’ 과제는 다자간 대화에서 발생하는 여러 주제를 식별하고 각 주제에 대한 요약문을 생성하는 것

자연어 생성 (NLG)

컴퓨터가 인간이 이해할 수 있는 형태의 텍스트를 자동으로 생성하는 과정

요약(Summarization)

중요한 문장을 뽑아내서 배열하는 **추출 요약**

텍스트를 이해하고 핵심적인 내용을 직접 생성해서 요약물을 작성하는 **추상 요약** ✓

01. 공모전 소개

데이터

이 과제의 데이터는 일상에서 이루어지는 대화를 대상으로 하며 ,
대화 내 각 주제의 핵심 내용을 효율적으로 추출하여 요약 정보를 제공한다 .

데이터 형식

주제별 대화	키워드	주제별 요약문
<div> 화자1: 우리 그때 처음에 나 데리고 가 줬던 초밥집 기억나? 화자2: Company-name 상암동에 있는 데? 화자1: 어 나 진짜 깜짝 놀랐잖아. 나는 화자2:그때까지 초밥은 그냥 회전 초밥집? 너가 데리고 가 줘서 오마카세 처음 알았어. 화자1: 그때 나는 애는 초밥이 뭔지 모르나 보더라는 생각이 들어서 데리고 갔지 화자2: 무시하는 거야? (...) </div>	초밥집	이 대화에서 화자 1과 2는 초밥집에 대해 이야기했습니다. 1은 2에게 자신을 데리고 가 줬던 초밥집에서 오마카세를 처음 알게 되었고 신세계였다고 이야기했습니다. 그리고 음식은 배 채우려고 먹는 게 아니라 맛있는 음식을 먹어야 한다고 생각했다고 이야기했습니다(...)

데이터 세트 구성

	훈련	검증	시험
자료 수	506	102	408

01. 공모전 소개

평가지표

아래 세 지표의 평균

루지-1 (ROUGE-1)

생성 텍스트와 참조 텍스트 사이의 일치하는 단어의 개수를 기반으로 평가

버트스코어 (BERTScore)

생성 텍스트와 참조 텍스트의 단어 임베딩 간의 유사도를 평가

블루트 (BLEURT)

사전 훈련된 언어모델을 통해 문장 쌍 간의 의미적 유사성을 평가



02. 모델 선정 과정

02. 모델 선정 과정

blossom -8b

코랩 프로 환경에서
학습불가

T5, Kobart v1

성능 개선 시도

KoGPT, KoBertSum
blossom 양자화

라이브러리 및 레퍼런스 부족

Kobart v2

baseline 모델로 선정

03. 모델 학습 과정 - 전처리

```
--
speaker: "SD2000002"
utterance: "어~ 혹시 포르투갈이나 스페인 유럽 쪽 다녀오신 적 있으신가요?"
▼ 6:
speaker: "SD2000001"
utterance: "어~ 네, 저도 우연히 스페인과 포르투갈을 다녀왔었습니다."
▼ 7:
speaker: "SD2000001"
▶ utterance: "어~ 저는 스페인 중에서도 마드리드에 근교에... 있구나라는 것을 새롭게 알게 됐었습니다."
▼ 8:
speaker: "SD2000001"
▶ utterance: "어~ 또 톨레도에 지역 음식도 같이 먹었었는데...약간 소박한 맛이 있었다고 생각을 합니다."
▼ 9:
speaker: "SD2000001"
▶ utterance: "어~ 또 물론 마드리드도 굉장히 좋았는데 유...고 해서 혹시 톨레도도 가본 적이 있나요?"
▼ 10:
speaker: "SD2000002"
▶ utterance: "아~ 제가 톨레도도 다녀왔는데 저는 이제 여...더웠기 때문에 많이 보진 못한 것 같아요."
▼ 11:
speaker: "SD2000002"
▶ utterance: "그때는 버스 관광버스를 타고 계속 돌아다니면...제 xx 장소로 넘어갔던 것 같 같습니다."
```

- 추임새 제거 (~)
- 불필요한 조사 제거 ("그", "이제", "막", "뭐" 등)
- 동어 반복 제거
- 동일 화자의 발화 연결

대화 생성 요약 시 필요한 정보는 전처리에서 제외



03. 모델 학습 과정

03. 모델 학습 과정 - 학습

파라미터 조정

- epochs: 전체 데이터셋에 대해 학습을 몇 번 반복할지를 설정
- batch_size: 학습 시 각 GPU 장치에서 사용되는 배치 크기
- learning_rate: 가중치 업데이트의 크기
 - lr_scheduler_type: 학습률 스케줄러의 유형을 설정
 - 코사인 함수 기반의 학습률 스케줄링등이 적용
 - warmup_steps: 학습률을 서서히 증가시키기 위해 설정된 스텝 수
 - 초기 학습 안정성 보장
- fp16 : 16비트 부동소수점 연산을 사용
 - 학습 속도를 높이고 메모리 사용량을 줄임

03. 모델 학습 과정 - 학습

1. PEFT 방법론

Parameter Efficient Fine Tuning

파인튜닝 과정에서 전체 파라미터를 학습하는 대신

task-specific한 파라미터만 학습하자

03. 모델 학습 과정 - 학습

모델 및 토크나이저 로드

```
model_name = "google/mt5-small"  
model = T5ForConditionalGeneration.from_pretrained(model_name)  
tokenizer = T5Tokenizer.from_pretrained(model_name)
```

PEFT 설정

```
lora_config = LoraConfig(  
    r=8,  
    lora_alpha=16,  
    target_modules=["q", "v"],  
    lora_dropout=0.1,  
    bias="none",  
)
```

← LoRA 적용

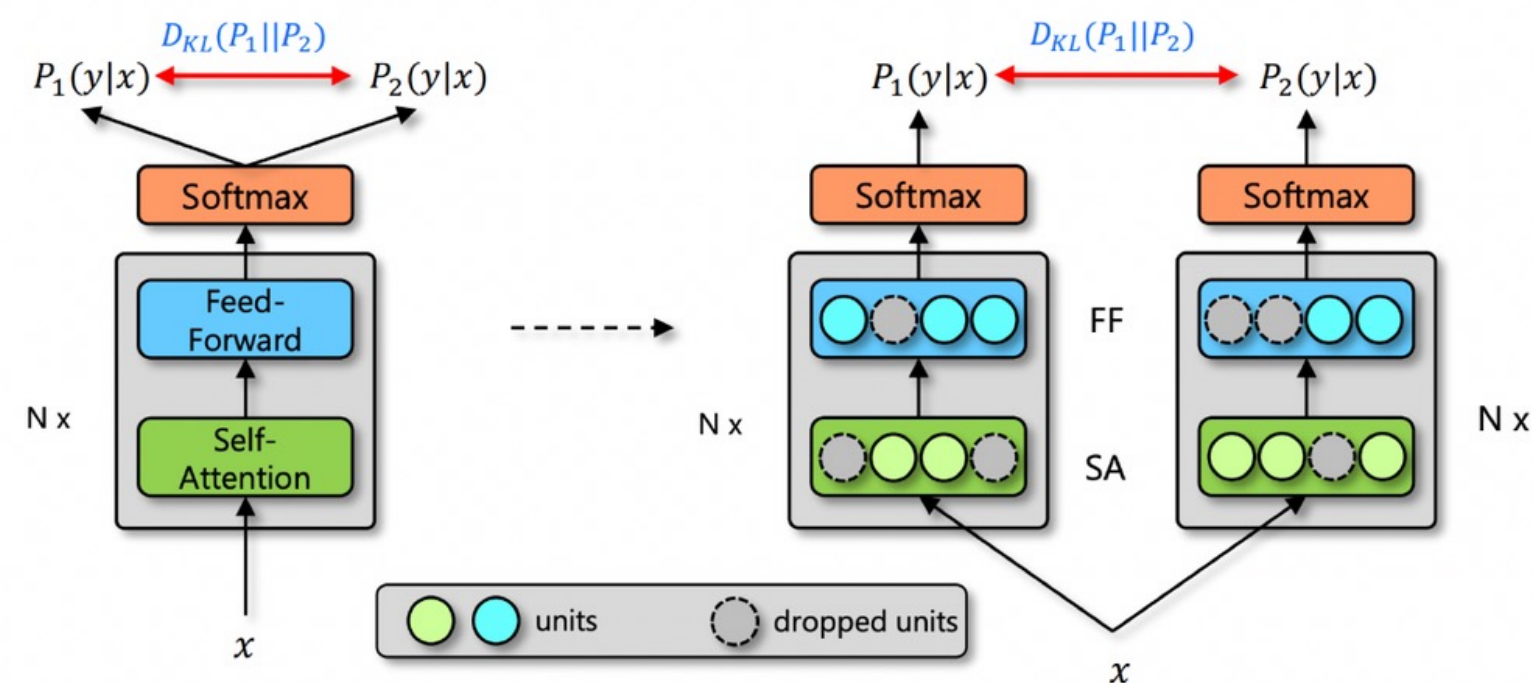
PEFT 적용

```
model = get_peft_model(model, lora_config)
```

03. 모델 학습 과정

2. Regularization

R-Drop(Reinforced Dropout)



드롭아웃(Dropout)을 두 번 다르게 적용하여,
동일한 입력에 대해 모델이 생성하는 두 개의 출력을 비교

두 분포 간의 Kullback-Leibler(KL) 발산을 최소화하도록 학습하는 방식

→ 결론적으로 다양한 드롭아웃 상태에서도 일관된 출력을 생성하게 해주고, 과적합을 방지함.

03. 학습

A) Kobart

- fine tuning 레퍼런스 모델
- 에폭 6이 가장 성능 높음

Step	Training Loss	Validation Loss
100	2.803900	2.253797
200	2.051700	1.842394
300	1.789500	1.723271

1.723

B) Kobart

- Kobart baseline 모델
- 전처리, 후처리,
파라미터 조정 시도

Step	Training Loss	Validation Loss
500	1.651000	2.430174
1000	1.335100	2.521549

2.521

C) T5

- T5 fine tuning
모델

Epoch	Training Loss	Validation Loss
1	16.514500	9.707318
2	7.220800	3.976629
3	4.106000	2.948559
4	3.724000	2.719298
5	3.373100	2.575845
6	3.332600	2.500924
7	3.010000	2.476337

D) T5

- T5 baseline 모델
- 전처리, 후처리,
파라미터 조정 시도

Step	Training Loss	Validation Loss
500	7.022300	7.029820
1000	6.667700	6.870688
1500	6.581400	6.883780
2000	6.333100	6.915590

Validation loss가 가장 낮은 KoBart 모델로 학습

03. 생성

탐색 알고리즘

- Greedy Search : 매 단계에서 가능한 다음 단어들 중 가장 높은 확률을 가진 단어를 선택
 - 빠르고 메모리 효율적이지만, 최적해를 찾기 어려울 수 있다.
- Beam. Search : 단일 경로가 아닌 여러 경로를 동시에 고려하여 탐색
 - 매 결정 단계에서 최적화가 이루어짐. 계산이 비교적 복잡하다.

파라미터 조정

- length penalty : 생성 문장 길이 조절
 - 증가시키면 문장 길이에 패널티 부여, 정보량 증가
- no_repeat_ngram_size : 생성 중 같은 N-그램을 반복하는 것을 허용
 - 일관성을 위해 반복 허용
- beam size : 탐색 과정에서 유지할 후보 경로의 수
 - 생성결과의 다양성과 비례함.
- Temperature : 무작위성, 창의성 조절
 - 낮을수록 보수적, 무작위성을 적절히 줄이면서 너무 경직되지 않도록 조정

03. 생성

A) Greedy Search

- max length 128
- early stopping : true

평가 점수	51.8230994	
과제명	지표명	평가 점수
일상 대화 요약	ROUGE-1	48.2052173
	bertscore	74.9833719
	bleurt	32.2807089

51.82

B) Beam search

- max length : 128
- beam size : 4
- no repeat ngram : 3
- early stopping : true
- length penalty : 1.0

평가 점수	51.7803181	
과제명	지표명	평가 점수
일상 대화 요약	ROUGE-1	48.0575496
	bertscore	74.8932509
	bleurt	32.390154

51.78

C) Beam search ★

- max length : 128
- beam size : 6
- no repeat ngram : 3
- early stopping : true
- length penalty : 0.9

평가 점수	52.3573209	
과제명	지표명	평가 점수
일상 대화 요약	ROUGE-1	47.851989
	bertscore	74.8248461
	bleurt	34.3951277

52.35

D) Beam search

- C 조건 동일
- top_p : 0.95
- top_k : 50
- 역변역

평가 점수	51.5173513	
과제명	지표명	평가 점수
일상 대화 요약	ROUGE-1	42.7030326
	bertscore	73.09413
	bleurt	38.7548912

51.51



04. 결과 기술

04. 결과

model	ROUGE-1	Bertscore	Bluert	score
Kobart greedy search	48.21	74.98	32.28	51.82
Kobart beam search beamsize: 4 / l.p: 10	48.06	74.89	32.39	51.78
Kobart beam search beamsize: 6 / l.p: 0.9	47.85	74.82	34.39	52.35 ⭐

04. 결과

"output": "두 화자는 이 대화에서 여름에 먹는 음식에 대해 말했습니다. SD2000038은 여름에는 팔빙수를 많이 먹는데 초코 빙수가 제일 맛있다고 말했고, 겨울에는 비앙코란 아이스크림을 먹었으며 가을에 먹어도 맛있는 거 같다고 이야기했습니다. 그리고 겨울에 먹을 수 있는 과일이 없어서 항상 아이스 음료만 먹게 된다고 이야기했고, 여름을 기다릴 때 선풍기와 에어컨 바람을 쐬면서 티브이를 볼 때 추억이 가장 큰 것 같고, 겨울을 따뜻하게 날 수 있기 때문에 항상 시원한 음료를 먹는다고 답했습니다. 또한 겨울에도 과일을 많이 먹는 편이라고 이야기했고 여름이 되면 복숭아와 수박을 먹으면서 티브를 볼 때의 즐거움이 제일 커서 여름은 추억을 제일 값진 기억으로 기억한다고 이야기하였습니다. 또한 여름의 가장 맛있는 음료는 단연 케이크를 곱았으며 과일은 비싸서 자주 먹지 않는다고 이야기했습니다, 007490은 겨울이 추우면 얼음이 얼어서 아이스 커피나 아이스로 된 음료가 가장 맛있고 가을에는 맛있을 거라는 이야기를 했습니다."

04. 결과

"output " : "두 화자는 이 대화에서 좋아하는 음식과 싫어하는 음식에 대해 말했습니다. SD2001182는 고깃집이나 패밀리 레스토랑에 가서 고기를 구워 먹거나 소주와 맥주를 마시는 편이라고 답했습니다. 또 가족들과 외식할 때는 주로 술자리를 선호한다고 말했고, 가족들이나 지인들 중에 가봤다는 곳이 있으면 가보고 싶다고 했습니다. 또 고기는 잘 못 먹지만, 야채와 과일을 많이 넣어서 먹는 편이고, 바삭바삭한 식감이 인상적이었다고 이야기했습니다. 이 밖에 음식은 술을 많이 먹지는 않지만, 가족이나 지인들과 함께 식사할 때 소주를 많이 마신다고 말했습니다, 가족들끼리 식사를 할 때 주로 술을 마시며, 술자리에서는 즐거운 분위기를 즐긴다고 덧붙였습니다."

04. 한계 및 느낀점

- PEFT 적용 실패로 더 큰 모델을 사용하지 못함
- 전처리 결과 성능 저하
- 발화자의 ID mapping 문제
- Bluert 점수 개선 어려움
 - 반복 문장 적은 생성 요약이 일관성을 낮춰 오히려 부정적인 영향을 미침

04. 한계 및 느낀점

• 발화자의 ID mapping 문제

- ▼ 1:
 - speaker: "SD2000164"
 - utterance: "결혼에 대한 가치관 이 생각이 어떠신지 물어봐도 될까요?"
- ▼ 2:
 - speaker: "SD2000165"
 - ▶ utterance: "사실 저는 지금 결혼을 별로 하고 싶지가 않...문에 ~ 지금은 그거에 대한 생각이 없지만"
- ▼ 3:
 - speaker: "SD2000165"
 - utterance: "전에 결혼을 하고 싶었을 때에는 저는 ~ 그냥 아기들과 잘 놀아주고"
- ▼ 4:
 - speaker: "SD2000165"
 - ▶ utterance: "가정에 충실한 가정에 충실하고 바람 피지 않...그런 화목한 가정을 너무 많이 바랐어가지고"
- ▼ 5:
 - speaker: "SD2000165"
 - utterance: "그냥 그런 남편을 많이 원했던 것 같아요."
- ▼ 6:
 - speaker: "SD2000164"
 - utterance: "저는 결혼을 추구해요."
- ▼ 7:
 - speaker: "SD2000164"
 - utterance: "결혼을 하고 싶은 마음을 지금도 가지고 있는데 ~ 사실 저에게는 조카가 조카들이 있어요."

▼ output:

"이 대화에서 화자들은 결혼에 대한 생각과 자녀 계획에 대해 이야기했습니다. 화자 SD2000961은 결혼을 별로 원하지 않아서 비혼주의자이기 때문에 결혼은 꼭 하고 싶다고 말했습니다. 또 상대방도 같이 성장하고 승승장구할 수 있게 도와주고 싶은 마음도 있다고 말했고, 결혼이 인생의 전부라고 생각하기 때문에 혼자 잘 살고 싶어서 커리어도 쌓으면서 커리어를 쌓을 생각이라고 덧붙였습니다. 한편 화자가 결혼 적령기인 20대 초반의 남자친구와 사귀어 봤는데 남자들은 항상 변하는 모습을 보면서 상처도 많이 받았고, 남자를 믿지 않게 되었던 것이 가장 큰 상처였다고 말하며 결혼 후 남편과 함께 살 수 있는 환경을 만드는 게 가장 중요하다는 것을 깨달았다고 이야기하면서 화자는 결혼을 통해 상대방과 함께 성장해 나갈 수 있도록 돕는 것이 결혼의 가장 중요한 부분이라고 생각한다고 했습니다. 또 화자의 나이가 어리긴 하지만 아직 어리지만 아직 나이는 많이 어리지만, 아직 남자들이 많이 사랑해주고 잘해 주던 사람들이 마지막에는 변해버리는 모습을 보며 많이 믿게 되었다고 말하며 화자도 결혼하고 싶지만 상대방을 믿으려고 노력하지 않는다고 말하며 자신은 결혼하지 않고 남편을 원하고 있다고 답했습니다."

실제 발화자와 요약된 결과에서의 ID가

일치하지 않는 문제점 발생

04. 한계 및 느낀점

- PEFT 적용 실패로 더 큰 모델을 사용하지 못함
- 전처리 결과 성능 저하
- 발화자의 ID mapping 문제
- Bluert 점수 개선 어려움
 - 반복 문장 적은 생성 요약이 일관성을 낮춰 오히려 부정적인 영향을 미침



Thank You