

Prefix-Tuning: Optimizing Continuous Prompts for Generation

arXiv:2101.00190 (cs)

- 박지영

목차

1. Introduction

2. Related Work

3. Fine-Tuning

4. Prefix-Tuning

5. Experimental Setup

6. Main Results

7. Intrinsic Evaluation

8. Discussion & Conclusion

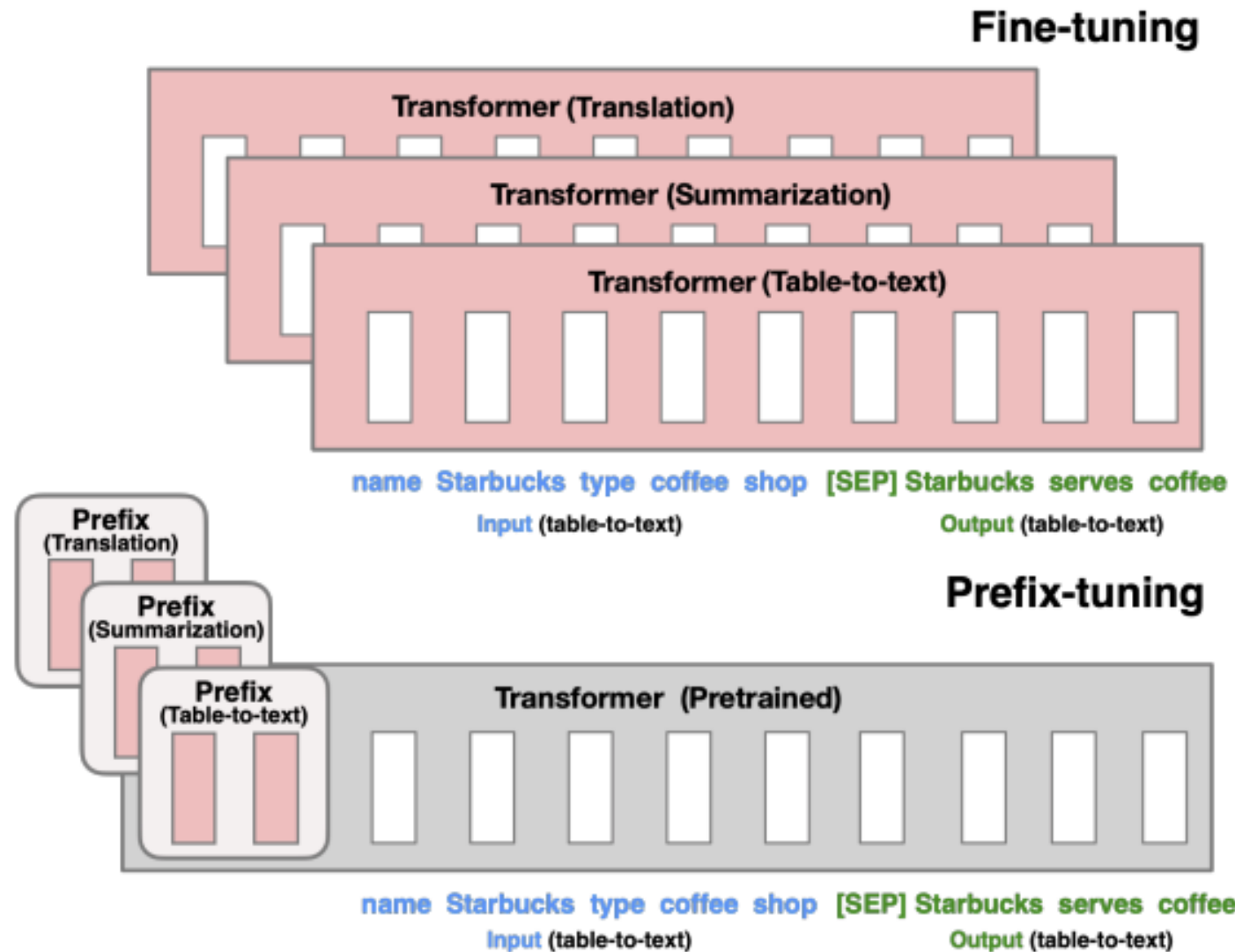
1. Introduction

Fine-tuning의 단점을 보완하고자
Prefix-tuning을 제안

Prefix 이용시,

Fine-tuning 같이 모델 전체를 학습
시키지 않고

Prefix부분만 조정하면 된다.



2. Related Work

- **Fine-tuning**

모델의 모든 매개변수를 특정 작업에 맞게 전체 조정

- **Lightweight fine-tuning**

대부분의 사전 훈련 매개변수 고정 > 일부 조정 (adapter layer 추가 및 조정)

- **Controllable generation.**

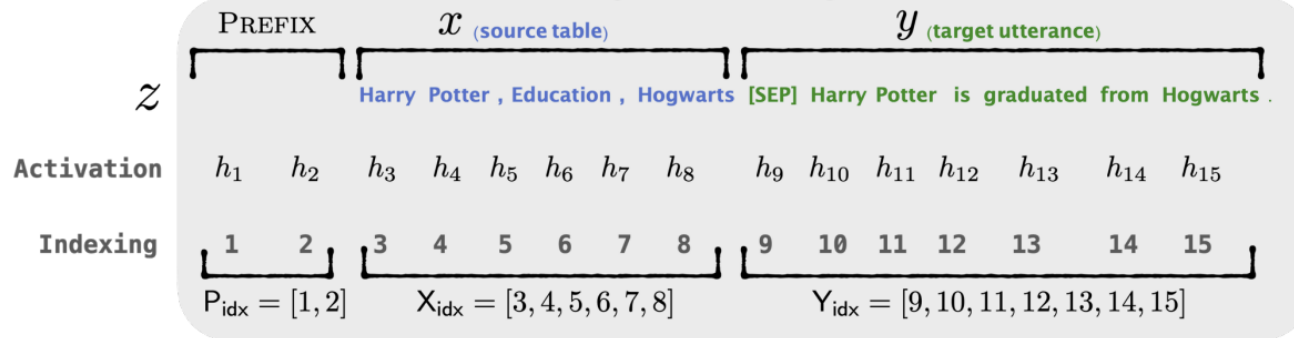
언어 모델의 출력을 특정 속성에 맞게 조정

- **Prompting**

지시사항, 예제(특정 트리거 단어 시퀀스)를 작업 입력 > 원하는 결과 출력

3. Fine-tuning

Autoregressive Model (e.g. GPT2)



Summarization Example

Article: Scientists at University College London discovered people tend to think that their hands are wider and their fingers are shorter than they truly are. They say the confusion may lie in the way the brain receives information from different parts of the body. Distorted perception may dominate in some people, leading to body image problems ... [ignoring 308 words] could be very motivating for people with eating disorders to know that there was a biological explanation for their experiences, rather than feeling it was their fault."

Summary: The brain naturally distorts body image - a finding which could explain eating disorders like anorexia, say experts.

Encoder-Decoder Model (e.g. BART)

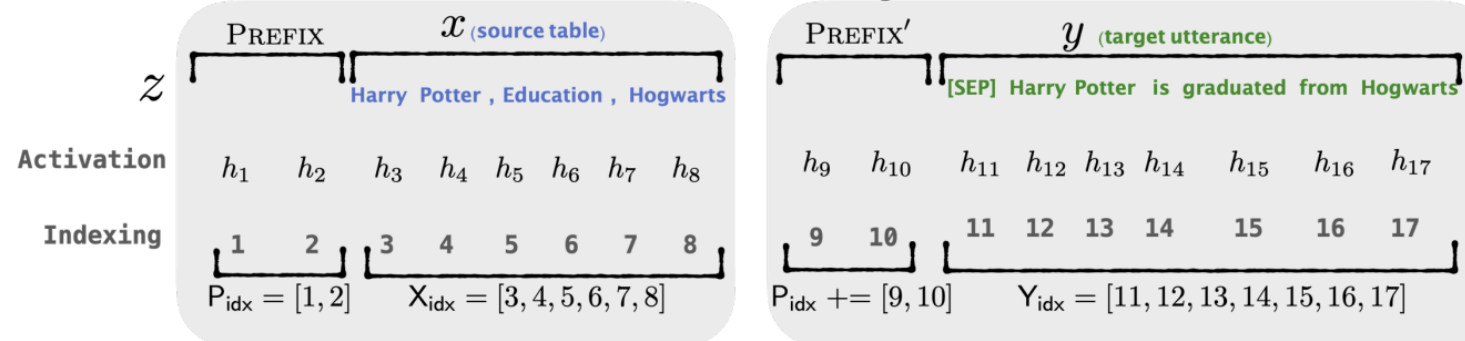


Table-to-text Example

Table: name[Clowns] customer-rating[1 out of 5] eatType[coffee shop] food[Chinese] area[riverside] near[Clare Hall]

Textual Description: Clowns is a coffee shop in the riverside area near Clare Hall that has a rating 1 out of 5 . They serve Chinese food .

3. Fine-tuning

- pretrained parameters ϕ > initialize
- p_ϕ = trainable language model distribution
- gradient updates (log-likelihood objective)

$$\max_{\phi} \log p_{\phi}(y \mid x) = \sum_{i \in Y_{\text{idx}}} \log p_{\phi}(z_i \mid h_{<i}).$$

4. Prefix-Tuning

collocation

Obama + Barack

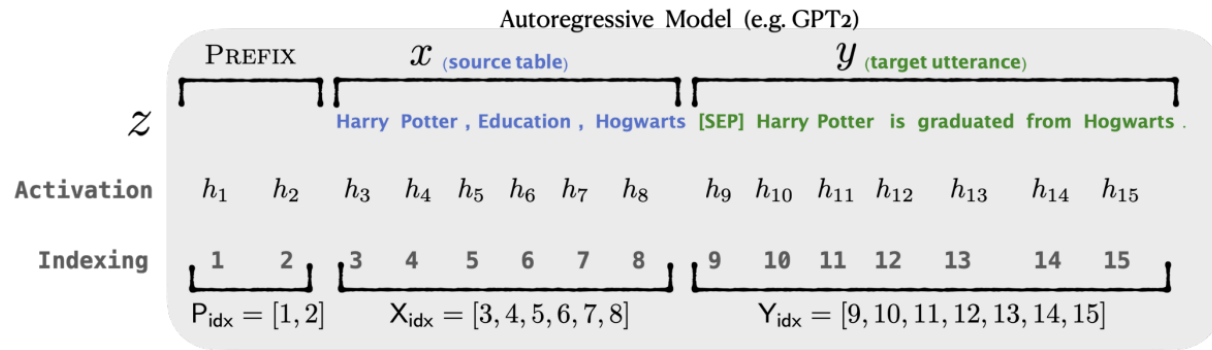
Instead of discrete tokens,

Sequential word embedding 최적화 가능

Pros and cons: expressive

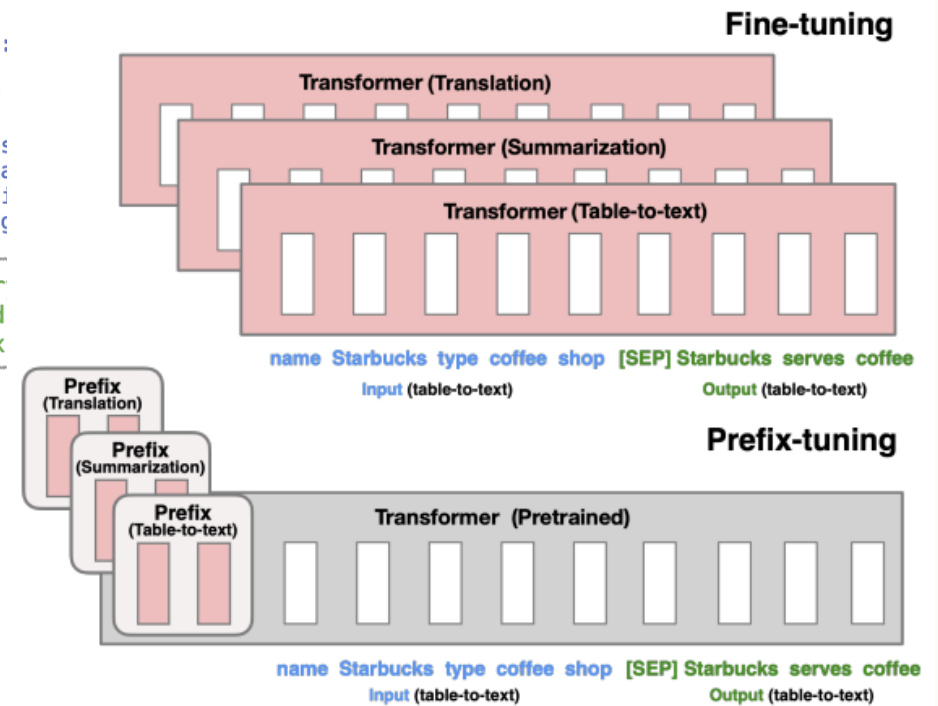
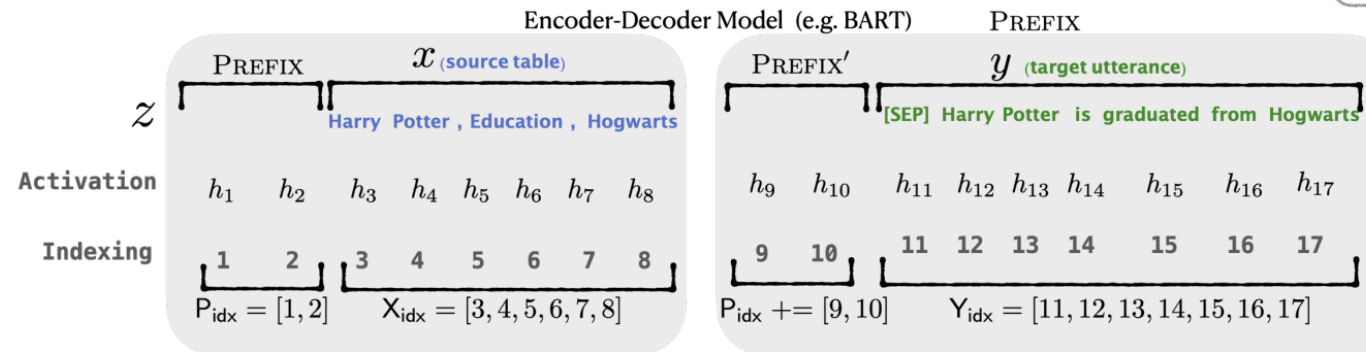
discrete tokens < Prefix-Tuning < all layers of the activations

4. Prefix-Tuning



Article:
tend to
shorter
way the
body.Dis
body ima
motivati
a biolog
feeling

Summar
a find
anorex



4. Prefix-tuning

- Parametrization of prefix parameters
- 문제점: **directly updating** > **unstable optimization**
- 해결책: **Reparametrize**

$$P_{\theta}[i, :] = \text{MLP}_{\theta}(P'_{\theta}[i, :]) \text{ by a smaller matrix } (P'_{\theta})$$

- 차원을 줄이면서 매개변수 감소 > 최적화 과정 안정화
- Mlp 통과하면서 동일한 차원으로 확장 > 복잡한 패턴 특성 학습 + 안정성

5. Experimental Setup

Experiment:

1. table-to-text : GPT-2 Medium, Large 모델
2. summarization : BART Large 모델

Data set:

table-to-text : E2E, WebNLG, DART
summarization : XSUM

비교 방법:

Table-to-text Example

FINE-TUNE
FT-TOP2
ADAPTER(3%)
ADAPTER(0.1%)
PREFIX(0.1%)

Summarization Example

FINE-TUNE
PREFIX

평가 매트릭스:

BLEU, NIST, METEOR, ROUGE-L, CIDEr

Summarization Example

Article: Scientists at University College London discovered people tend to think that their hands are wider and their fingers are shorter than they truly are. They say the confusion may lie in the way the brain receives information from different parts of the body. Distorted perception may dominate in some people, leading to body image problems ... [ignoring 308 words] could be very motivating for people with eating disorders to know that there was a biological explanation for their experiences, rather than feeling it was their fault."

Summary: The brain naturally distorts body image - a finding which could explain eating disorders like anorexia, say experts.

Table-to-text Example

Table: name[Clowns] customer-rating[1 out of 5] eatType[coffee shop] food[Chinese] area[riverside] near[Clare Hall]

Textual Description: Clowns is a coffee shop in the riverside area near Clare Hall that has a rating 1 out of 5 . They serve Chinese food .

6. Main Results

Table-to-text Example

	E2E					WebNLG									DART					
	BLEU	NIST	MET	R-L	CIDEr	BLEU			MET			TER ↓			BLEU	MET	TER ↓	Mover	BERT	BLEURT
						S	U	A	S	U	A	S	U	A						
GPT-2 _{MEDIUM}																				
FINE-TUNE	68.2	8.62	46.2	71.0	2.47	64.2	27.7	46.5	0.45	0.30	0.38	0.33	0.76	0.53	46.2	0.39	0.46	0.50	0.94	0.39
FT-TOP2	68.1	8.59	46.0	70.8	2.41	53.6	18.9	36.0	0.38	0.23	0.31	0.49	0.99	0.72	41.0	0.34	0.56	0.43	0.93	0.21
ADAPTER(3%)	68.9	8.71	46.1	71.3	2.47	60.4	48.3	54.9	0.43	0.38	0.41	0.35	0.45	0.39	45.2	0.38	0.46	0.50	0.94	0.39
ADAPTER(0.1%)	66.3	8.41	45.0	69.8	2.40	54.5	45.1	50.2	0.39	0.36	0.38	0.40	0.46	0.43	42.4	0.36	0.48	0.47	0.94	0.33
PREFIX(0.1%)	69.7	8.81	46.1	71.4	2.49	62.9	45.6	55.1	0.44	0.38	0.41	0.35	0.49	0.41	46.4	0.38	0.46	0.50	0.94	0.39
GPT-2 _{LARGE}																				
FINE-TUNE	68.5	8.78	46.0	69.9	2.45	65.3	43.1	55.5	0.46	0.38	0.42	0.33	0.53	0.42	47.0	0.39	0.46	0.51	0.94	0.40
Prefix	70.3	8.85	46.2	71.7	2.47	63.4	47.7	56.3	0.45	0.39	0.42	0.34	0.48	0.40	46.7	0.39	0.45	0.51	0.94	0.40
SOTA	68.6	8.70	45.3	70.8	2.37	63.9	52.8	57.1	0.46	0.41	0.44	-	-	-	-	-	-	-	-	-

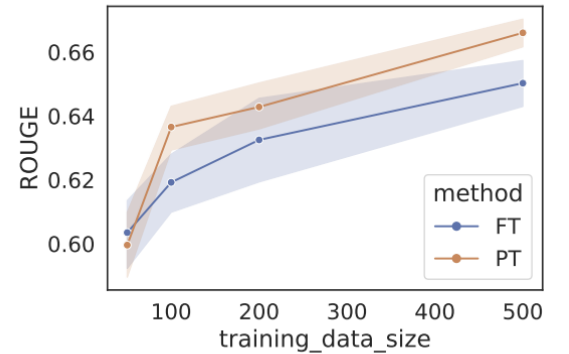
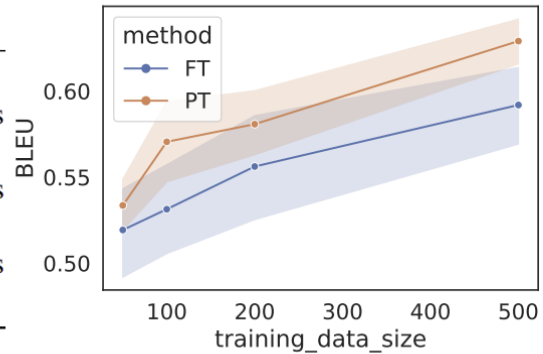
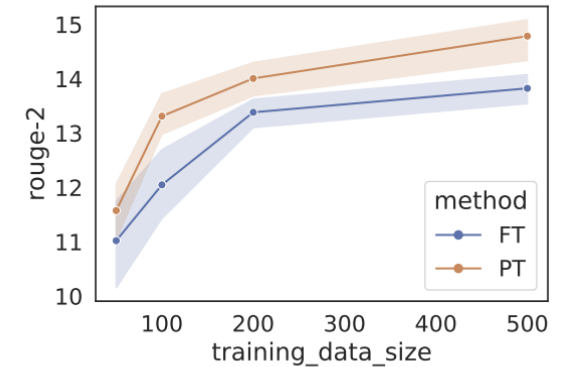
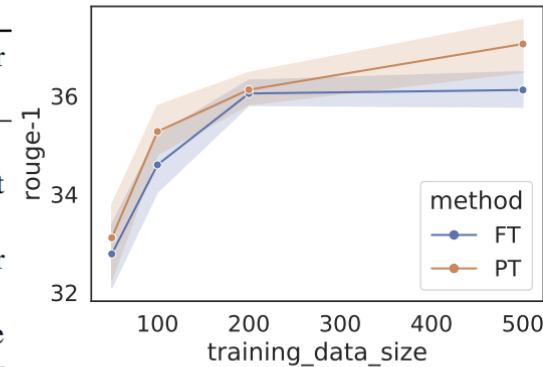
Summarization Example

	R-1 ↑	R-2 ↑	R-L ↑
FINE-TUNE(Lewis et al., 2020)	45.14	22.27	37.25
PREFIX(2%)	43.80	20.93	36.05
PREFIX(0.1%)	42.92	20.03	35.05

6. Main Results

Low-data Setting

Source	name : The Eagle type : coffee shop food : Chinese price : cheap customer rating : average area : riverside family friendly : no near : Burger King
Prefix (50)	The Eagle is a cheap Chinese coffee shop located near Burger King.
Prefix (100)	The Eagle is a cheap coffee shop located in the riverside near Burger King. It has average customer ratings.
Prefix (200)	The Eagle is a cheap Chinese coffee shop located in the riverside area near Burger King. It has average customer ratings.
Prefix (500)	The Eagle is a coffee shop that serves Chinese food. It is located in the riverside area near Burger King. It has an average customer rating and is not family friendly.
FT (50)	The Eagle coffee shop is located in the riverside area near Burger King.
FT (100)	The Eagle is a cheap coffee shop near Burger King in the riverside area. It has a low customer rating and is not family friendly.
FT (200)	The Eagle is a cheap Chinese coffee shop with a low customer rating. It is located near Burger King in the riverside area.
FT (500)	The Eagle is a cheap Chinese coffee shop with average customer ratings. It is located in the riverside area near Burger King.



6. Main Results

Extrapolation : 훈련 데이터에 포함 되지 않은 데이터에 대한 예측

Table-to-text Example

	BLEU			WebNLG			TER ↓		
	S	U	A	S	U	A	S	U	A
GPT-2 _{MEDIUM}									
FINE-TUNE	64.2	27.7	46.5	0.45	0.30	0.38	0.33	0.76	0.53
FT-TOP2	53.6	18.9	36.0	0.38	0.23	0.31	0.49	0.99	0.72
ADAPTER(3%)	60.4	48.3	54.9	0.43	0.38	0.41	0.35	0.45	0.39
ADAPTER(0.1%)	54.5	45.1	50.2	0.39	0.36	0.38	0.40	0.46	0.43
PREFIX(0.1%)	62.9	45.6	55.1	0.44	0.38	0.41	0.35	0.49	0.41
GPT-2 _{LARGE}									
FINE-TUNE	65.3	43.1	55.5	0.46	0.38	0.42	0.33	0.53	0.42
Prefix	63.4	47.7	56.3	0.45	0.39	0.42	0.34	0.48	0.40
SOTA	63.9	52.8	57.1	0.46	0.41	0.44	-	-	-

Summarization Example

	news-to-sports			within-news		
	R-1 ↑	R-2 ↑	R-L ↑	R-1 ↑	R-2 ↑	R-L ↑
FINE-TUNE	38.15	15.51	30.26	39.20	16.35	31.15
PREFIX	39.23	16.74	31.51	39.41	16.87	31.47

7. Intrinsic Evaluation

1. Prefix length

Summarization Example

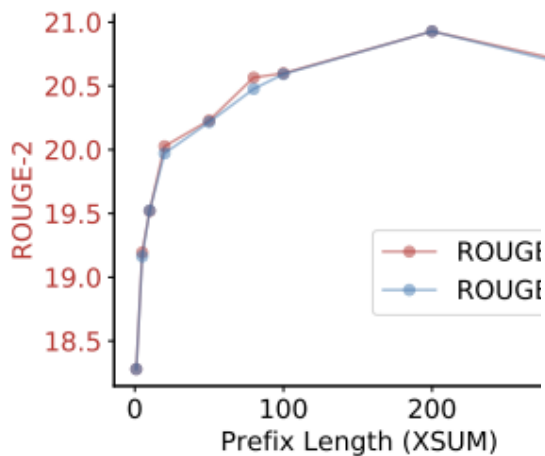
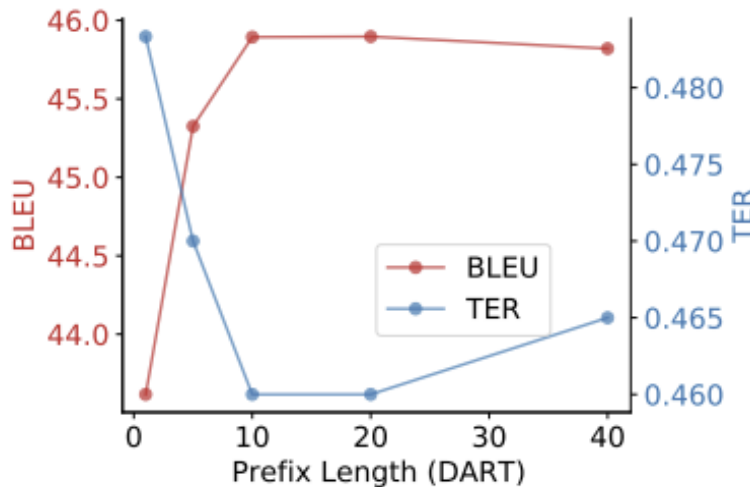
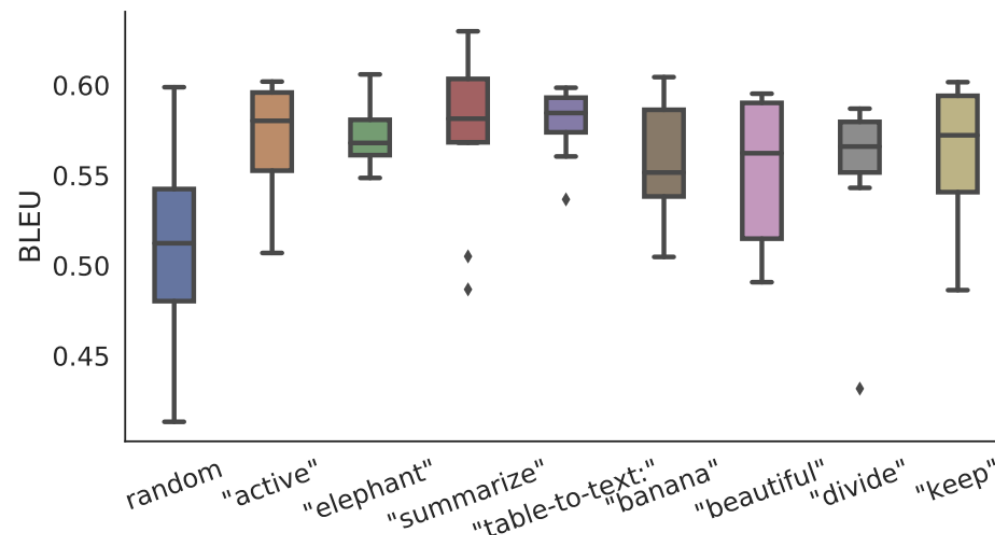


Table-to-text Example



2. prefix 매개변수 초기화



8. Discussion & Conclusion

Prefix-tuning 제안

- fine-tuning과 비슷하거나 더 높은 성능
- prefix 조정 모듈 및 공간 효율적

개인 정보 보호와 같은 실용적인 방안

- Personalization & Batching Across Users

일반화 능력 향상

- Inductive Bias of Prefix-tuning

Thank you

