

### I. 서론

#### 연구 목적

GTP-3, BART, BERT 등과 같은 대규모 사전 학습된 언어 모델을 활용하기 위해 일반적으로 Fine-tuning을 수행하나 이는 언어 모델의 기존 매개변수를 저장하여 메모리 부담과 더불어 방대한 GPU 자원을 요구하는 단점을 보인다. 따라서 언어 모델의 크기가 점점 증대되는 추세에 맞추어 효율적이고 경량화된 모델 활용 방법이 주목받고 있으며 본 연구에서는 한국어 자연언어처리 태스크를 위하여 가상 토큰인 Prefix를 추가하여 Pre-trained Prefix Prompting 모델을 구축하고 한국어 자연언어처리 태스크에 활용하는 방안을 제시한다.

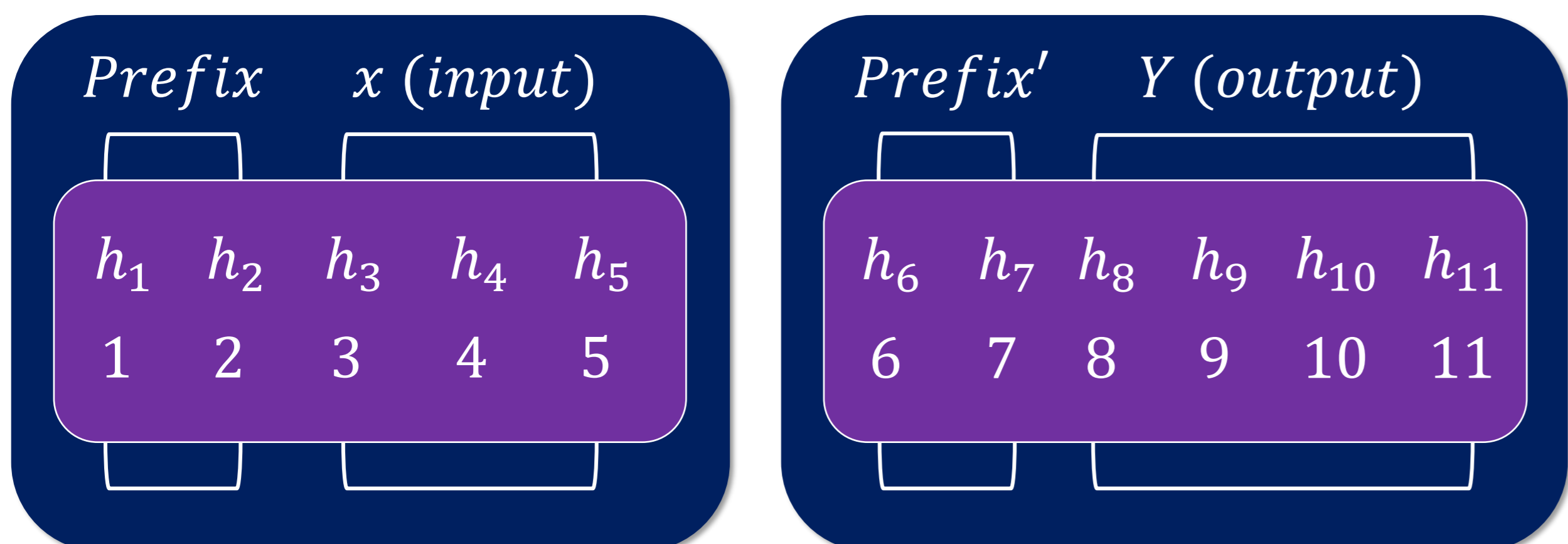
### II. 실험 설계

#### 모델 구성

$$h_i = \begin{cases} P_\theta[i:] & \text{if } (i \in P_{idx}) \\ LM_\phi(z_i, h < i), & \text{otherwise.} \end{cases}$$

기존 논문에서 자연언어 생성 태스크를 위하여 그림 [1]과 같이 인코더와 디코더 모두에  $z = [Prefix; x; Prefix'; y]$  형태로 Prefix를 추가하였으나 본 논문에서는 자연언어 감성 분류 태스크에 대해 적용하기 위해 그림 [2]와 같이 인코더에만  $z = [Prefix; x]$  형태로 Prefix를 추가한 형태로 모델을 구성한다.  $P_{idx}$ 는 추가된 Prefix를 나타내며  $|P_{idx}|$ 는 Prefix의 길이로 정의할 때, 모델의 activation  $h_i$ 는 위 수식과 같다.

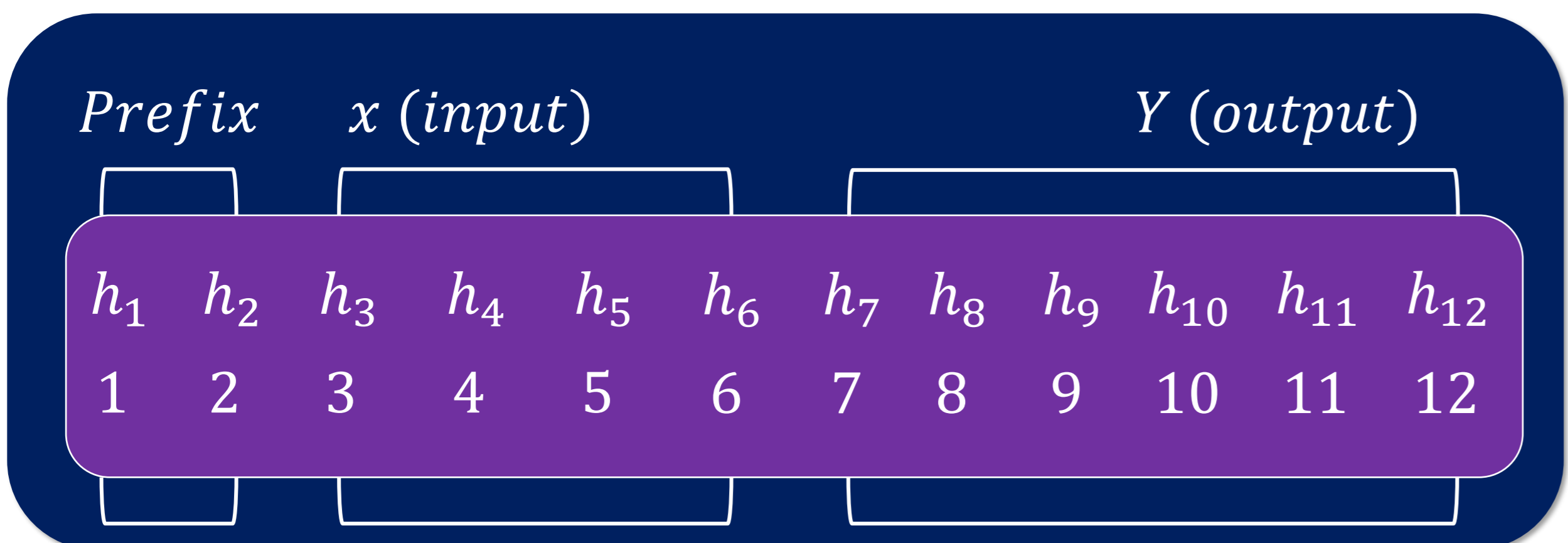
Encoder – Decoder Model



$P_{idx} = [1, 2]$        $P_{idx} += [6, 7]$   
 $X_{idx} = [3, 4, 5]$        $Y_{idx} = [8, 9, 10, 11]$

그림 1: Prefix-tuning 인코더-디코더 모델 구조

Autoregressive Model



$P_{idx} = [1, 2]$        $X_{idx} = [3, 4, 5, 6]$        $Y_{idx} = [7, 8, 9, 10, 11, 12]$

그림 2: 한국어 자연언어처리 분류 태스크를 위한 모델 구조

KLUE-RoBERTa 모델에 위와 같이 Prefix Prompt를 추가한 뒤 Masked Language Modeling 방식으로 학습시켜 모델의 기존 매개변수는 고정하고 Prefix Prompt를 조정하는 매개변수만 최적화를 수행한다. 이후 Prefix Prompting 모델에 기반하여 한국어 감성 분류 태스크를 수행하고 성능을 측정한다.

#### 데이터 셋

Prefix Prompting 학습에 활용된 한국어 위키피디아 데이터 셋은 한국어 위키피디아 문서의 약 2700만여 문장을 제공하고 있으며 각각 0.5%의 테스트 데이터와 개발 데이터를 포함한다.

모델 성능 평가에 활용된 네이버 영화리뷰 데이터 셋은 총 20만 개의 한국어 영화리뷰 문장으로 이루어져 있고 각 문장에 긍정 및 부정의 이진 레이블이 포함된다. 긍정 데이터와 부정 데이터는 균형 있게 구성되었으며 테스트 데이터로 5만 문장을 포함한다.

### III. 실험 결과

표 1: 한국어 자연언어 감성 분류 예측 정확도

Model	Accuracy
RoBERTa	62.97%
Prefix-Length10	61.85%
Prefix-Length50	68.31%
Prefix-Length100	64.55%
Prefix-Length200	67.97%
RoBERTa (Fine-tuning)	90.78%
Prefix-Length10 (Fine-tuning)	90.83%
Prefix-Length50 (Fine-tuning)	90.89%
Prefix-Length100 (Fine-tuning)	90.88%
Prefix-Length200 (Fine-tuning)	90.68%

#### 결과 분석

표 [1]은 사전 학습 언어 모델의 매개변수를 고정한 RoBERTa 및 Prefix Prompting 기반 분류 모델의 정확도를 측정한 결과와 해당 모델을 Fine-tuning하여 정확도를 측정한 결과이다. Prefix Prompting 모델은 사전 학습 시 기존 120M여 개에 달하는 RoBERTa 모델의 전체 매개변수 대비 약 8%인 9.8M의 매개변수만을 새롭게 학습한다.

Prefix Prompting 모델의 정확도 성능이 증가함을 보이는데 매개변수를 고정한 경우 Prefix Prompting 모델의 정확도가  $|P_{idx}|$ 에 따라 최대 약 7%p 향상되었으며 이는 RoBERTa 모델을 기반으로 Fine-tuning 수행한 결과에 비해 약 22%p의 정확도 차이를 보여 Prefix Prompting 모델을 이용한 경우에 Fine-tuning 대비 성능 차이를 줄일 수 있음을 보인다.

또한 Prefix Prompting 모델을 기반으로 Fine-tuning을 수행한 실험에서도 기존 모델 대비 정확도가 최대 0.11%p 상승함을 보여 적절한 Prefix 길이를 설정하고 Prefix Prompting 모델을 구성한 경우 보편적인 성능 향상을 보장함을 보인다.

### IV. 결론

#### 결론

Prefix Prompting 모델은 사전 학습 과정에서 약 8%의 매개변수만을 새롭게 학습함으로써 학습 시간이나 메모리 용량 측면에서 이점을 보이면서도 Prefix 길이에 따라 메모리 용량 대비 성능을 조절할 수 있어 기존 모델을 기반으로 Fine-tuning을 수행한 결과에 근접한 성능을 보장한다. 또한 Prefix Prompting 모델을 기반으로 Fine-tuning을 수행한 태스크에서도 소폭의 성능 향상을 보여 경량화 Fine-tuning을 요구하는 환경에서 Prefix Prompting 모델을 구성한 뒤 필요에 따라 모델의 성능을 증강시킬 수 있어 보편적으로 다양한 학습 환경에 적용할 수 있는 강점을 보인다.