Artficial Intelligence Laboratory

Neural Grapheme-to-Phoneme Conversion with Pre-trained Grapheme Models

2022 IEEE

University of Science and Technology of China,Hefei
iFLYTEK Research

정보융합공학과 AI전공
정주경

부산대학교
PUSAN NATIONAL UNIVERSITY

목 차

1. Introduction

2. Proposed Method

3. Experiments

4. Results

5. Conclusion

Introduction

G2P

- 초기: joint n-gram model, joint sequence model, WFST
- 최근: neural networks(LSTM, Transformer)

GBERT(Grapheme BERT)

- Transformer 기반 G2P 모델 개선
- Multi-layer Transformer encoder, grapheme만 있는 데이터로 self-supervised 사전 학습
- Masked grapheme 예측

GBERT Transformer 개선

1. GBERT fine-tuning
2. GBERT를 Transformer model에 attention하여 fused(융합) => BERT-fused[1]

[1] Incorporating BERT into Neural Machine Translation

Proposed Method

Grapheme BERT(GBERT)

- Multi-layer bidirectional Transformer encoder
- GBERT: 한 단어의 grapheme sequence vs BERT: 한,두 문장 wordpiece sequence
- Single-word G2P task 초점(단일 단어 grapheme sequence만 고려)

- Masked grapheme 예측 작업을 통해 사전 학습

=> input grapheme 일부 랜덤 masking하고 masked grapheme 예측

- Mask token(80%), random grapheme(10%), original grapheme(10%)

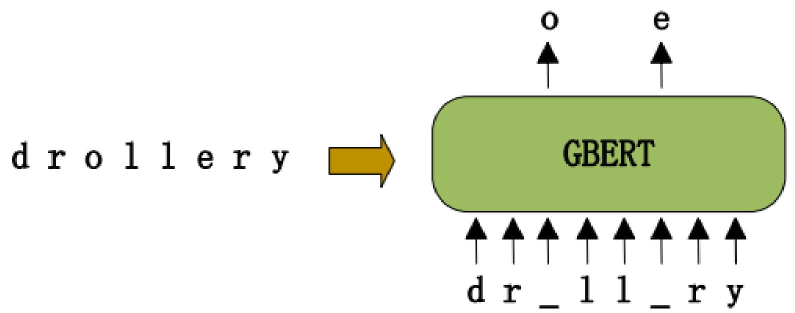


Fig. 1. The masked grapheme prediction task for pre-training GBERT. Here is an example of the English word *drollery*. The "_" denotes the mask token.

Proposed Method

Fine-tuning GBERT for G2P conversion

- Vanila Transformer encoder를 GBERT로 대체하고 end-to-end로 학습
- 사전 학습된 인코더와 randomly initialized 디코더에 서로 다른 learning rate 사용하면 더 나은 수렴[1]

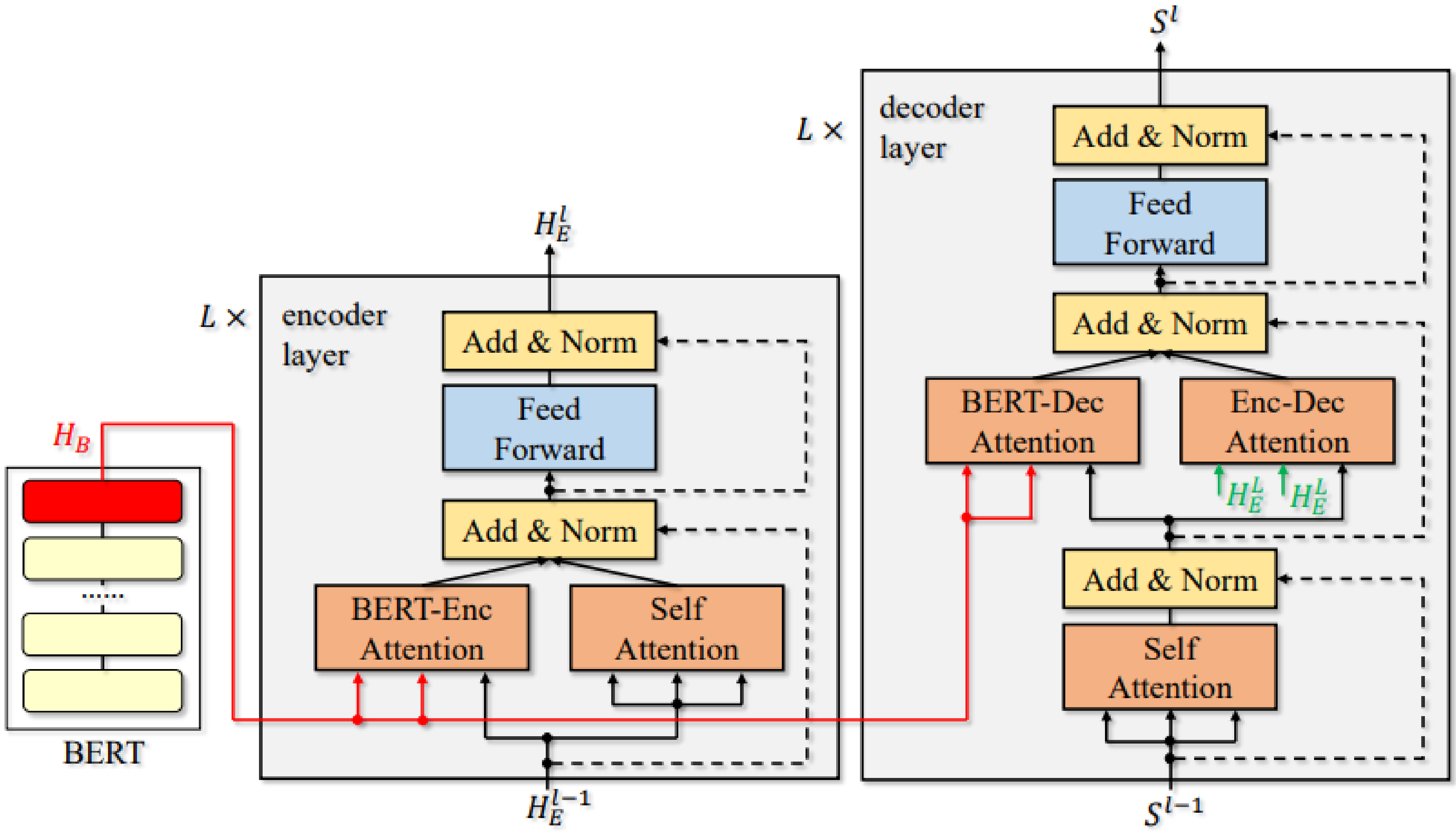
Fusing GBERT into Transformer-based G2P model

- 사전 학습된 언어 모델을 통합하는 또 다른 방법은 feature extractor
- BERT-fused: vanila Transformer 인코더, 디코더 레이어와 BERT output feature 상호 작용

⇒ medium-resource 기계 번역에서 fine-tuning보다 더 잘 작동

[1] Yang Liu and Mirella Lapata, “Text summarization with pretrained encoders,”

BERT-fused


S^l
$L \times$
decoder
layer
Add & Norm
Feed
Forward
Add & Norm
BERT-Dec
Attention
Enc-Dec
Attention
H_E^L
H_E^L
Add & Norm
Self
Attention
S^{l-1}
H_E^l
$L \times$
encoder
layer
Add & Norm
Feed
Forward
Add & Norm
BERT-Enc
Attention
Self
Attention
H_E^{l-1}
H_B
BERT

Proposed Method

Fusing GBERT into Transformer-based G2P model

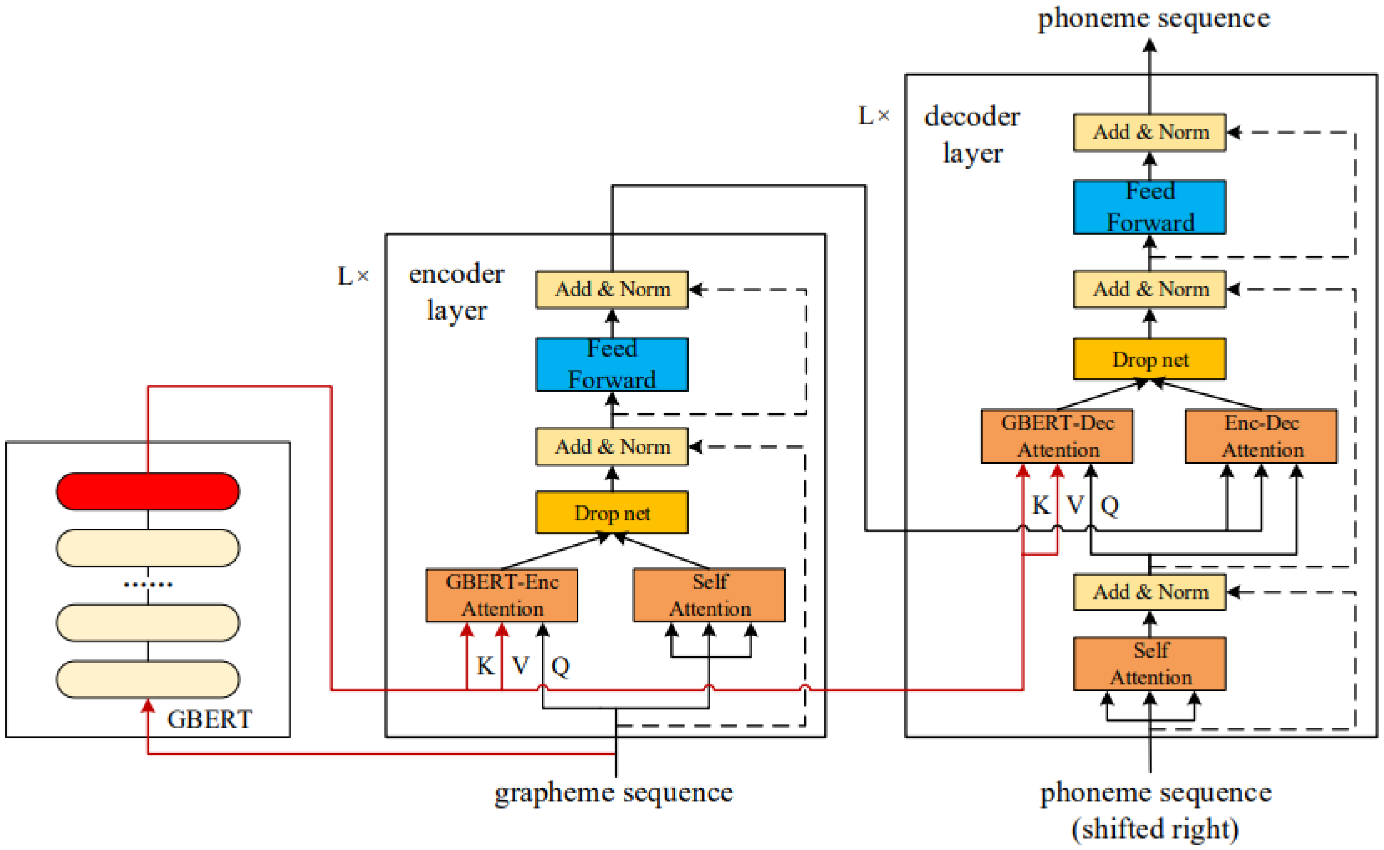


Fig. 2. The architecture of GBERT-fused model, which fuses GBERT into the Transformer-based G2P model and is adapted from [16].

Experiments

G2P Dataset

- SIGMORPHONE 2021 G2P task dataset
- Medium-resource 10개 언어 중 가장 어려운 4개 언어 선택
- 학습(8,000), 검증(1,000), 테스트(1,000)
- 한글 자모 사용(가감 -> ㄱ ㅏ ㄱ ㅏ ㅁ)
- Low-resource: 학습 셋에서 1000개 샘플링

Table 1. The four languages used in our experiments.

Language	Language Family	Script Type	Word Example
Dutch	Germanic	Latin	afnemer
Serbo-Croatian	South Slavic	Latin	buđenje
Bulgarian	East Slavic	Cyrillic	абоната
Korean	Koreanic	Hangul	가치관

GBERT Dataset

- GBERT 사전 학습 위해 WikiPron에서 데이터 수집
- 검증 및 테스트셋 단어 제외
- 각 언어 90% 학습, 10% 검증

GBERT 데이터

네덜란드어	세르보크로아티아	불가리아어	한국어
27,000	35,000	43,100	14,100

Experiments

Implementation

- Masked grapheme ratio: 20%
- 한 단어의 평균 grapheme 수가 평균 단어 수 보다 작음

⇒ 낮은 비율은 모델이 grapheme간의 문맥적 관계 발견X

Table 3. The influence of the masked grapheme ratio for pre-training GBERT on the performance of the GBERT attention model for the medium-resource Dutch G2P task.

Mask Ratio (%)	Mask Accuracy (%)	WER (%)	PER (%)
15	53.01	16.34 ± 0.43	3.51 ± 0.08
20	53.48	15.86 ± 0.21	3.38 ± 0.07
30	51.32	15.88 ± 0.30	3.40 ± 0.08

Masked grapheme 예측 정확도

네덜란드어	세르보크로아티아	불가리아어	한국어
53.48%	58.43%	80.66%	40.63%

Experiments

Model

Imitation Learning	SIGMORPHON 2021 모방 학습 단일 모델
Transformer	Transformer 기반 아키텍처
GBERT w/o fine-tuning	GBERT 파라미터 고정(frozen)
GBERT fine-tuning	Pre-trained GBERT encoder + Transformer decoder
GBERT attention	BERT-fused기반

Results

Table 2. The WER and PER results (%) of different models on our medium-resource and low-resource G2P tasks.

Model	Dutch		Serbo-Croatian		Bulgarian		Korean	
	WER	PER	WER	PER	WER	PER	WER	PER
medium-resource								
IL [21]	17.7 ± 1.3	-	38.9 ±1.2	-	19.7 ± 1.7	-	18.9 ± 0.8	-
Transformer	16.86 ± 0.14	3.49 ± 0.05	39.50 ± 0.71	7.94 ± 0.14	20.72 ± 2.70	3.84 ± 0.74	19.26 ± 0.48	3.39 ± 0.10
GBERT w/o fine-tuning	19.98 ± 0.66	4.61 ± 0.48	43.92 ± 0.97	9.53 ± 0.75	23.24 ± 2.54	4.17 ± 0.40	23.06 ± 0.30	4.92 ± 0.15
GBERT fine-tuning	16.18 ± 0.23	3.59 ± 0.39	39.08 ± 1.05	7.83 ± 0.21	18.88 ± 2.31	**3.42 ± 0.33**	19.80 ± 0.62	3.52 ± 0.12
GBERT attention	**15.86 ± 0.21**	**3.38 ± 0.07**	**37.64 ± 0.71**	**7.67 ± 0.20**	**18.60 ± 1.92**	**3.42 ± 0.34**	**17.94 ± 0.71**	**3.16 ± 0.13**
low-resource								
Transformer	34.30 ± 0.66	9.39 ± 0.42	68.86 ± 1.08	15.41 ± 0.16	33.06 ± 1.90	6.07 ± 0.35	**29.72 ± 0.77**	**6.79 ± 0.35**
GBERT w/o fine-tuning	34.56 ± 0.72	9.13 ± 0.60	69.70 ± 0.69	15.88 ± 0.65	41.87 ± 0.89	9.15 ± 0.88	42.76 ± 1.04	12.66 ± 0.29
GBERT fine-tuning	**28.96 ± 0.69**	**6.94 ± 0.61**	**63.12 ± 0.76**	**13.29 ± 0.57**	**30.86 ± 1.66**	**5.30 ± 0.36**	32.78 ± 1.21	8.43 ± 0.11
GBERT attention	35.28 ± 0.66	9.41 ± 0.33	68.14 ± 0.68	15.06 ± 0.29	31.98 ± 1.32	5.76 ± 0.35	29.78 ± 0.60	6.81 ± 0.30

한국어 성능

- **마스크 예측 정확도가 낮음 =>** 한국어의 grapheme간 문맥 관계가 약함

Results

Transfer Learning

- 저자원 성능 향상 위해 고자원 언어 같이 사용하는 GBERT기반 transfer learning 실험
- 저자원 언어: 네덜란드어, 고자원 언어:영어 => 게르만어족, 라틴어 문자 유형
- 사전 학습 데이터: WikiPron의 49,100개 영어 + 1,000개 네덜란드어

Table 4. transfer learning

Model	WER	PER
Transformer	24.54 ± 0.53	5.09 ± 0.22
GBERT w/o fine-tuning	33.30 ± 0.89	7.89 ± 0.25
GBERT fine-tuning	23.42 ± 0.82	4.84 ± 0.16
GBERT attention	**23.36 ± 0.82**	**4.76 ± 0.25**

Model	Dutch	
	WER	PER
medium-resource		
IL [21]	17.7 ± 1.3	-
Transformer	16.86 ± 0.14	3.49 ± 0.05
GBERT w/o fine-tuning	19.98 ± 0.66	4.61 ± 0.48
GBERT fine-tuning	16.18 ± 0.23	3.59 ± 0.39
GBERT attention	**15.86 ± 0.21**	**3.38 ± 0.07**
low-resource		
Transformer	34.30 ± 0.66	9.39 ± 0.42
GBERT w/o fine-tuning	34.56 ± 0.72	9.13 ± 0.60
GBERT fine-tuning	**28.96 ± 0.69**	**6.94 ± 0.61**
GBERT attention	35.28 ± 0.66	9.41 ± 0.33

Table 2. Dutch result

Conclusion

- **Pre-trained grapheme model GBERT 제안**
- **Transformer기반 G2P model 향상 위해 GBERT fine-tuning, GBERT attention**
- **GBERT attention 중간자원에서 모두 효과적**
- **GBERT fine-tuning 저자원에서 대부분 언어 효과적**
- **transfer learning을 통해 저자원 성능 향상**