

구글 BERT의 정석

다른 언어에 BERT 적용하기

20180376 안제준

M-BERT 이해하기

M-BERT

MLM과 NSP를 사용하면서 영어 위키피디아뿐 아니라 104개 언어의 위키피디아 텍스트로 학습

BERT

영어 위키피디아와 토론토 책 말뭉치를 가지고 마스크 언어 모델링(MLM)과 다음 문장 예측(NSP)로 학습

NLI 태스크 평가로 알아보는 M-BERT 평가하기

Language	Premise / Hypothesis	Genre	Label
English	You don't have to stay there. You can leave.	Face-To-Face	Entailment
French	La figure 4 montre la courbe d'offre des services de partage de travaux. Les services de partage de travaux ont une offre variable.	Government	Entailment
Spanish	Y se estremeció con el recuerdo. El pensamiento sobre el acontecimiento hizo su estremecimiento.	Fiction	Entailment
German	Während der Depression war es die ärmste Gegend, kurz vor dem Hungertod. Die Weltwirtschaftskrise dauerte mehr als zehn Jahre an.	Travel	Neutral
Swahili	Ni silaha ya plastiki ya moja kwa moja inayopiga risasi. Inadumu zaidi kuliko silaha ya chuma.	Telephone	Neutral
Russian	И мы занимаемся этим уже на протяжении 85 лет. Мы только начали этим заниматься.	Letters	Contradiction
Chinese	让我告诉你，美国人最终如何看待你作为独立顾问的表现。 美国人完全不知道您是独立律师。	Slate	Contradiction
Arabic	تحتاج الوكالات لأن تكون قادرة على قياس مستويات النجاح. لا يمكن للوكالات أن تعرف ما إذا كانت ناجحة أم لا	Nine-Eleven	Contradiction

Table 1: Examples (premise and hypothesis) from various languages and genres from the XNLI corpus.

XNLI dataset

장르	전제	가설	레이블
편지	여러분의 꿈을 우리의 꿈에 추가 할 것인가?	전국 최고의 학교를 세울 수 있도록 도와줄 수 있는가?	중립
911	구조 업무를 위해 뉴욕 소방국 (FDNY) 앤서니 부서장의 보고서를 살펴봐라.	앤서니 부서장은 구조 업무를 위한 보고서를 작성했다.	함의
소설	브리지타운에서 그날 밤 스페인 공습이 있었다.	브리지타운에서 공습이 일어나지 않았다.	모순
여행	말레콘 뒷골목에는 세련된 독특한 클럽들이 생겨나고 있다.	수집품이 증가하고 있지만 작년 만큼 빠르게 증가하지는 않는다.	중립
축어	중범죄는 줄었지만 살인사건은 증가했다.	살인사건은 증가하고 있다.	함의

그림 7-1 MLNLI 데이터셋

MNLI dataset

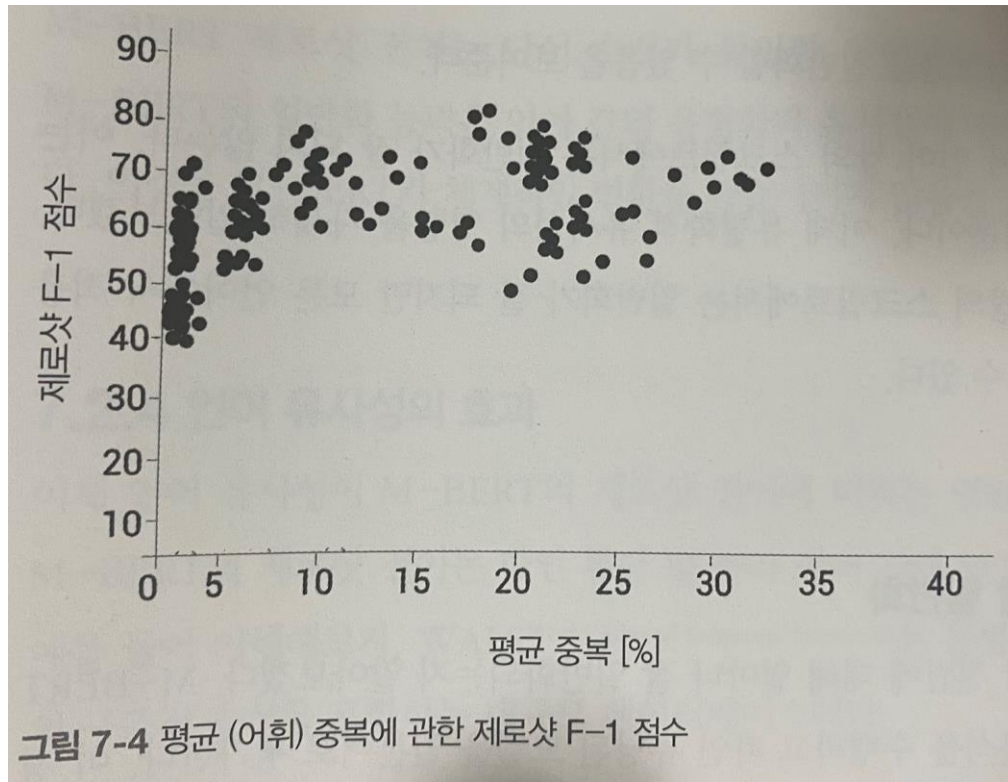
M-BERT를 평가하는 법

- 제로샷
 - 파인튜닝 : 영어 학습셋
 - 평가 : 모든 언어 테스트셋
- 번역-테스트
 - 파인튜닝 : 영어 학습셋
 - 평가 : 영어로 번역된 테스트셋
- 번역-학습
 - 파인튜닝 : 영어에서 다른 언어로 번역된 학습셋
 - 평가 : 모든 언어 테스트셋
- 번역-학습-모두
 - 파인튜닝 : 영어에서 다른 언어로 번역된 학습셋
 - 평가 : 모든 언어 테스트셋

M-BERT는 다국어 표현이 어떻게 가능한가?

모델	설정	영어	중국어	스페인어	독일어	아랍어	우르드어
BERT-cased	TRANSLATE -TRAIN	81.9	76.6	77.8	75.9	70.7	61.6
BERT-uncased	TRANSLATE -TRAIN	81.4	74.2	77.3	75.2	70.5	61.7
BERT-uncased	TRANSLATE -TEST	81.4	70.1	74.9	74.4	70.4	62.1
BERT-uncased	Zero-shot	81.4	74.3	74.3	70.3	60.1	58.3

어휘 중복 효과



스크립트에 대한 일반화

- 스크립트에 대한 일반화 : 서로 다른 언어에도 일반화해 적용할 수 있는가
 - 형태소 분석 태스크를 실험으로 우르두어를 이용해 M-BERT를 파인 튜닝
 - 그 후 힌디어로 M-BERT를 평가하면 91.1%의 평가 정확도를 보인다.

유형학적 특징에 대한 일반화

평가 언어 파인 튜닝 언어	영어	불가리아어	일본어
영어	96.8	87.1	49.4
불가리아어	82.2	98.9	51.6
일본어	57.4	67.2	96.5

언어 유사성의 효과

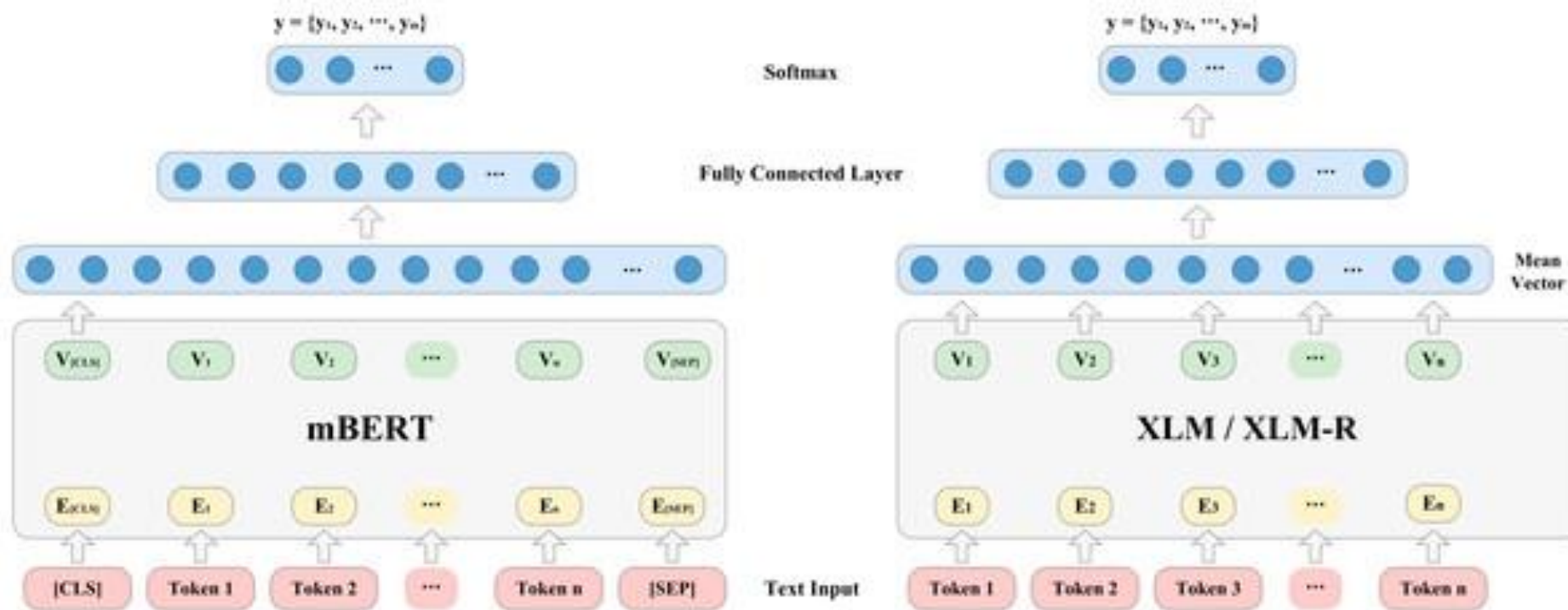
- WALS 공통 특징 수가 많을수록 제로샷 정확도가 높다
 - WALS : 문법, 어휘, 음운 속성 같은 언어의 구조적 속성을 포함한 대규모 데이터베이스

코드 스위칭과 음차의 효과

- 코드 스위칭 : 대화에서 서로 다른 언어를 혼합하거나 교대로 사용하는 것
 - ex) 영어 문장 사이에 한국어를 포함시키기(코리안은 굉장히 kind하고 polite한 것 같아)
- 음차 : 소스 언어 스크립트로 텍스트를 작성하지 않고 타깃 언어 스크립트를 사용
 - ex) 영어문장 사이에 포함된 한국어를 영어발음으로 작성(코리안은 굉장히

파인 튜닝셋	평가 셋	정확도
음차	음차	85.64
단일 언어(영어 + 힌디어)	음차	50.41
코드 스위칭	코드 스위칭	90.56
단일언어(영어 + 힌디어)	코드 스위칭	86.59

XLM



사전 학습 전략

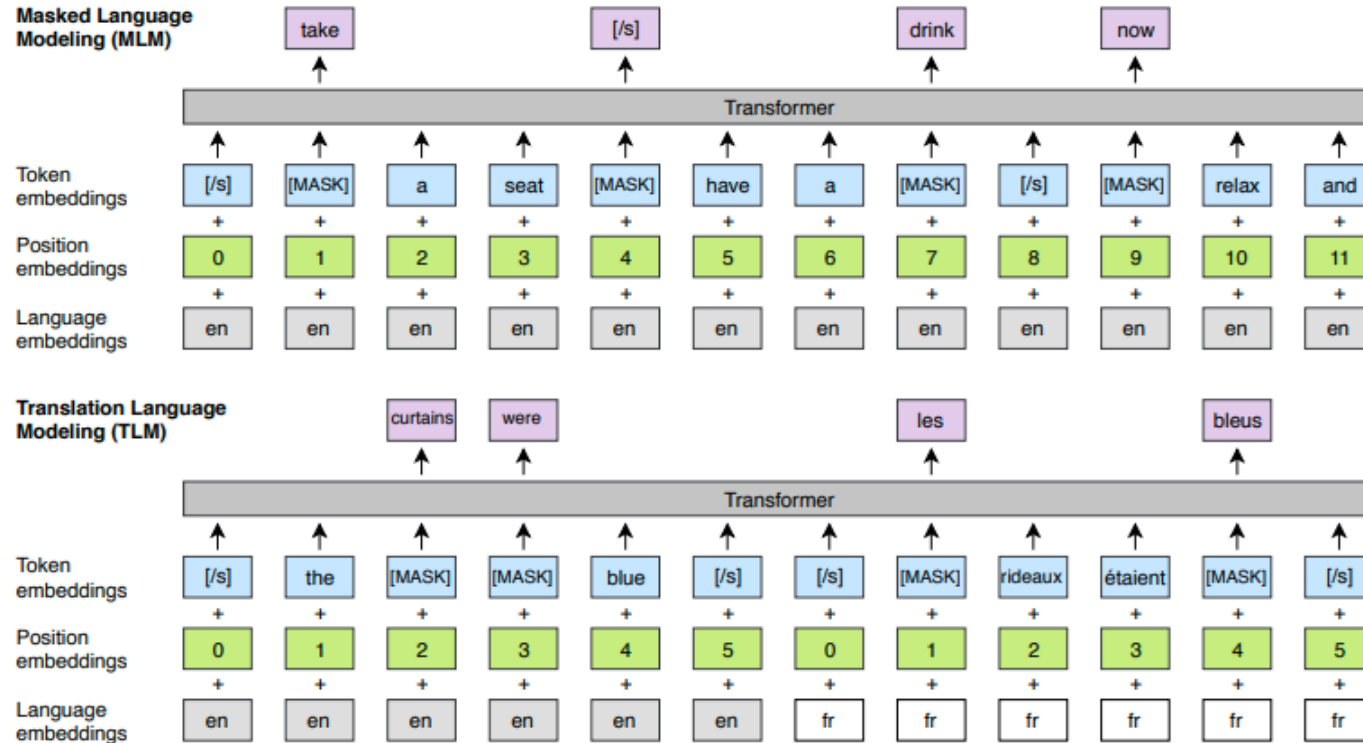


Figure 1: Cross-lingual language model pretraining. The MLM objective is similar to the one of [Devlin et al. \(2018\)](#), but with continuous streams of text as opposed to sentence pairs. The TLM objective extends MLM to pairs of parallel sentences. To predict a masked English word, the model can attend to both the English sentence and its French translation, and is encouraged to align English and French representations. Position embeddings of the target sentence are reset to facilitate the alignment.

XLM 사전 학습

- CLM사용
- MLM사용
- TLM과 결합해 MLM 사용

XLM 평가

모델	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	△
XLM(MLM)	83.2	76.5	76.3	74.2	73.1	74.0	73.1	67.8	68.5	71.2	69.2	71.9	65.7	64.6	63.4	71.5
XLM (MLM+TLM)	85.0	78.7	78.9	77.8	76.6	77.4	75.3	72.5	73.1	76.1	73.2	76.5	69.6	68.4	67.3	75.1

그림 7-11 XLM 평가

또한 연구원들은 [그림 7-12]와 같이 번역-학습 및 번역-테스트 환경에서 XLM을 평가했다. XLM이 M-BERT보다 더 좋은 성능을 보여준다.

모델	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	△
TRANSLATE-TRAIN																
M-BERT	81.9	-	77.8	75.9	-	-	-	-	70.7	-	-	76.6	-	-	61.6	-
XLM(MLM+TLM)	85.0	80.2	80.8	80.3	78.1	79.3	78.1	74.7	76.5	76.6	75.5	78.6	72.3	70.9	63.2	76.7
TRANSLATE-TEST																
M-BERT	81.4	-	74.9	74.4	-	-	-	-	70.4	-	-	70.1	-	-	62.1	-
XLM(MLM+TLM)	85.0	79.0	79.5	78.1	77.8	77.6	75.5	73.7	73.7	70.8	70.4	73.6	69.0	64.7	65.1	74.2

그림 7-12 XLM 평가

언어별 BERT

- 프랑스어 - FlauBERT
- 스페인어 - BETO
- 네덜란드어 -BERTje
- 독일어, 중국어, 일본어 - BERT
- 러시아어 - RuBERT