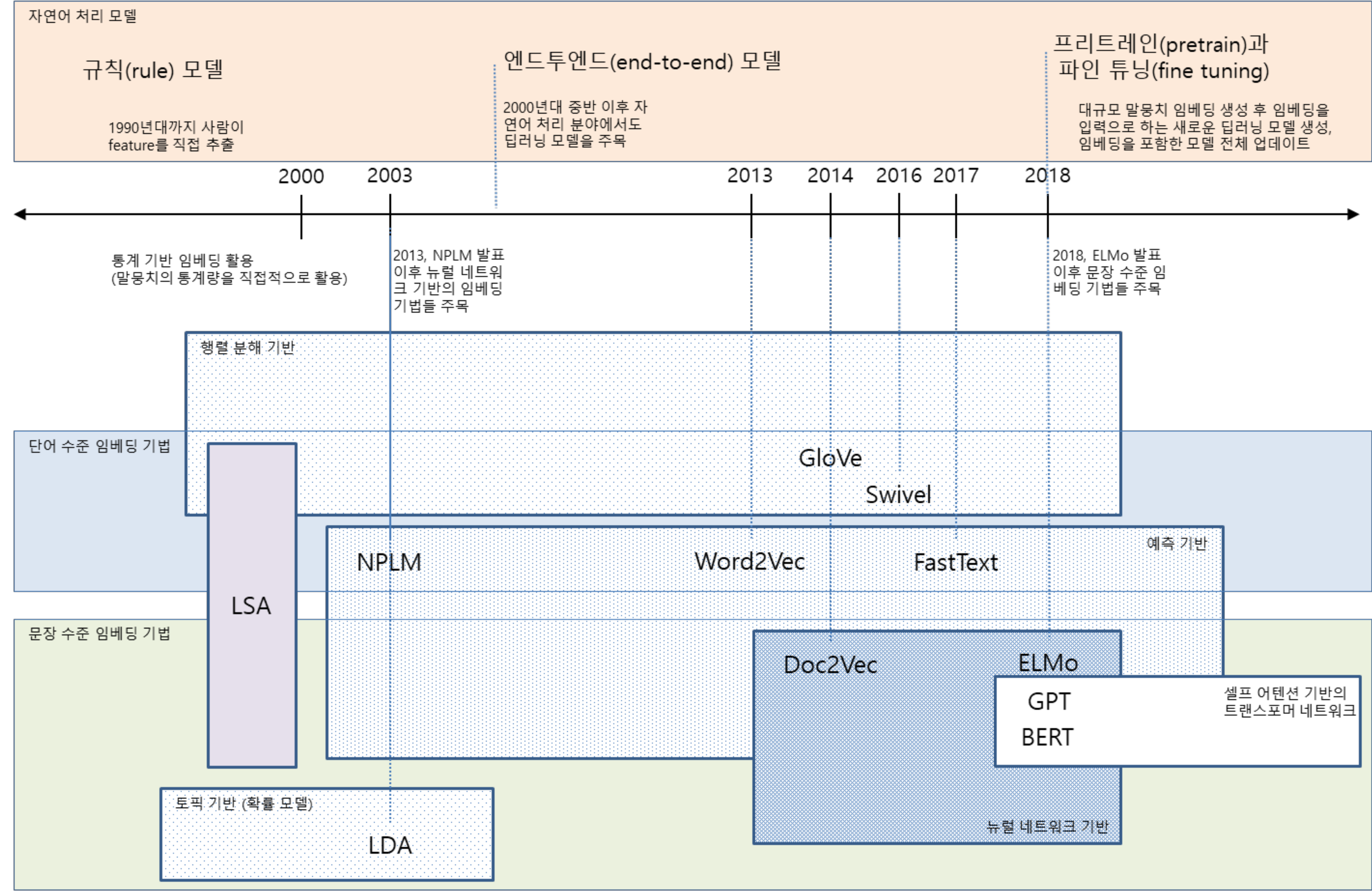


언어 모델 과거, 현재 그리고 미래

언어 모델의 개념과, BERT가 탄생하기 이전의 언어모델들을 살펴보겠습니다



Models

- word2vec, Glove, fastText,...
- Semi-supervised Sequence Learning - 2015, Google
- Deep contextualized word representations (ELMo) - 2018, Allen AI
- Universal Language Model Fine-tuning for Text Classification (ULMFiT) - 2018, fast.ai
- Improving Language Understanding by Generative Pre-Training - 2018, OpenAI
- BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding - 2018, Google AI Language
- Many more to come ... etc...

지금의 언어 모델

BERT 모델의 개념과 학습 방식에 대해 살펴보겠습니다

Bi-directional Encoder Representations from Transformers 5

BERT

- BERT는 bi-directional Transformer로 이루어진 언어모델
- 잘 만들어진 BERT 언어모델 위에 1개의 classification layer만 부착하여 다양한 NLP task를 수행
- 영어권에서 11개의 NLP task에 대해 state-of-the-art (SOTA) 달성

"Jimi" Hendrix was an American rock guitarist, singer, and songwriter.

1 classification layer for Fine-tuning

Pre-trained BERT

....
James Marshall "Jimi" Hendrix
was an American rock guitarist,
singer, and songwriter.
....

Who is Jimi Hendrix?

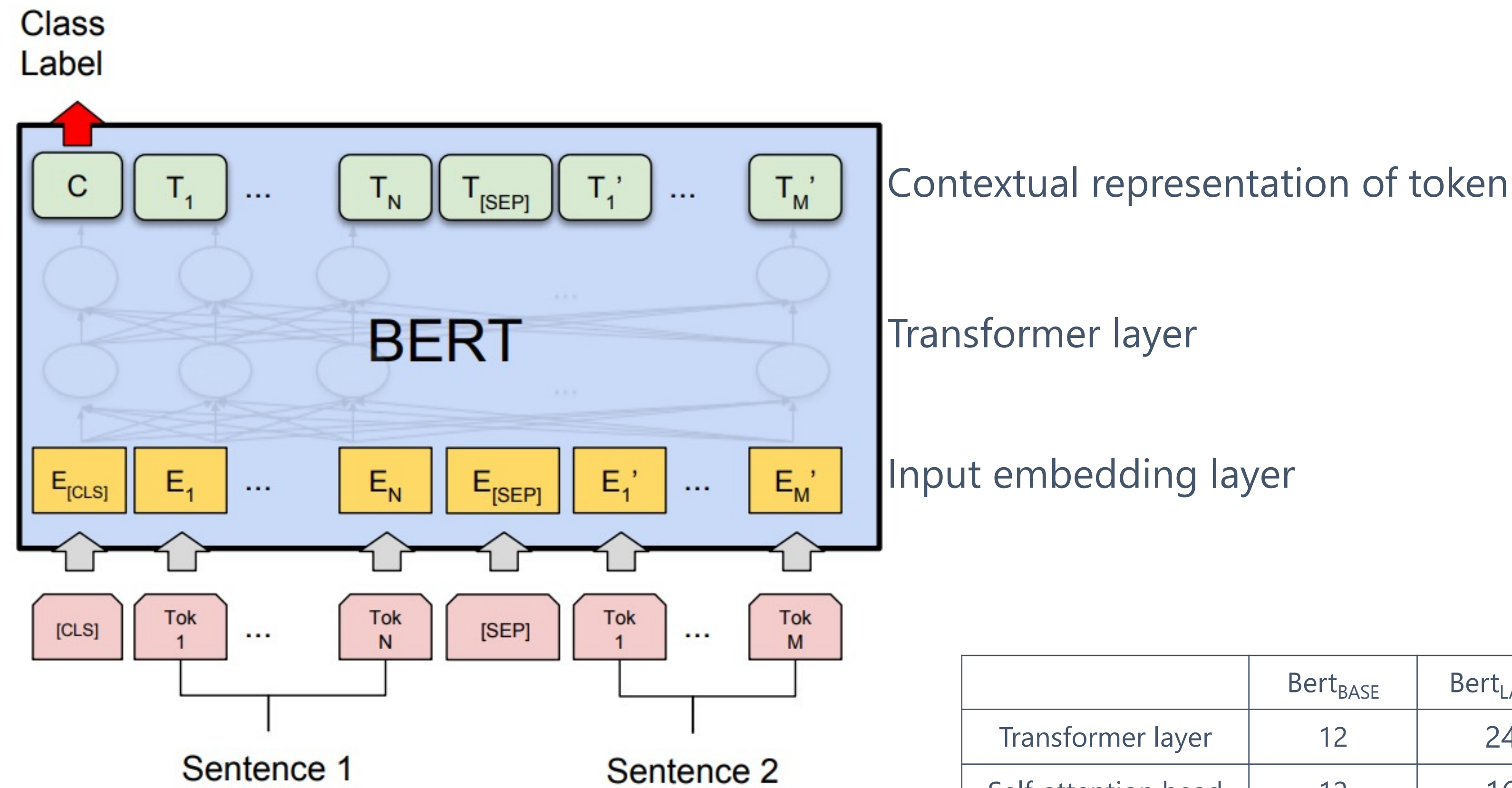
SQuAD v1.1 dataset leaderboard

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805	87.433	93.160
2 Oct 05, 2018	BERT (single model) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805	85.083	91.835

Methods – BERT 모델의 구조도

BERT

- Model architecture



	Bert _{BASE}	Bert _{LARGE}
Transformer layer	12	24
Self-attention head	12	16
Total	110M	340M

Methods – BERT 모델 학습 데이터

BERT

• 학습 코퍼스 데이터

- BooksCorpus (800M words)
- English Wikipedia (2,500M words without lists, tables and headers)
- 30,000 token vocabulary

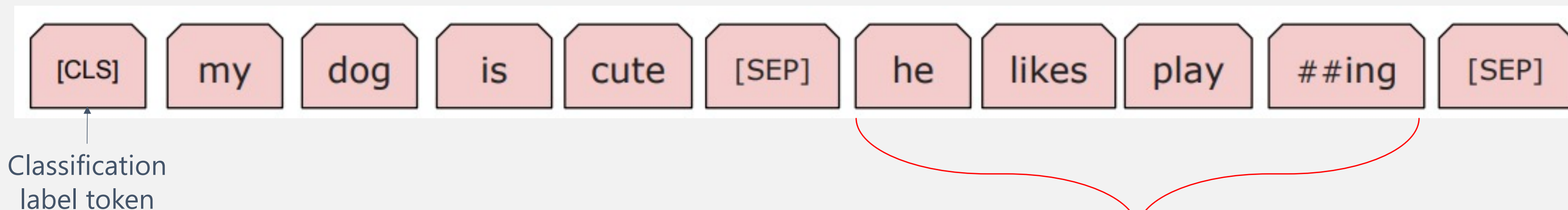
• 데이터의 tokenizing

- **WordPiece** tokenizing

He likes playing → He likes play **##ing**

- 입력 문장을 tokenizing하고, 그 token들로 'token sequence'를 만들어 학습에 사용
- 2개의 token sequence가 학습에 사용

Example of two sentences token sequence



Next sentence or Random chosen sentence (50%)

Methods – BERT의 WordPiece tokenizing

BERT

- Byte Pair Encoding (BPE) 알고리즘 이용
- 빈도수에 기반해 단어를 의미 있는 패턴(Subword)으로 잘라서 tokenizing

W2V vocabs

고양경찰서
고양시
종로경찰서
경찰
경찰서
경찰관
경찰청
⋮



고양	##경찰	##서
	##시	
종로	##경찰	##서
	경찰	
	경찰	##서
	경찰	##관
	경찰	##청
	⋮	



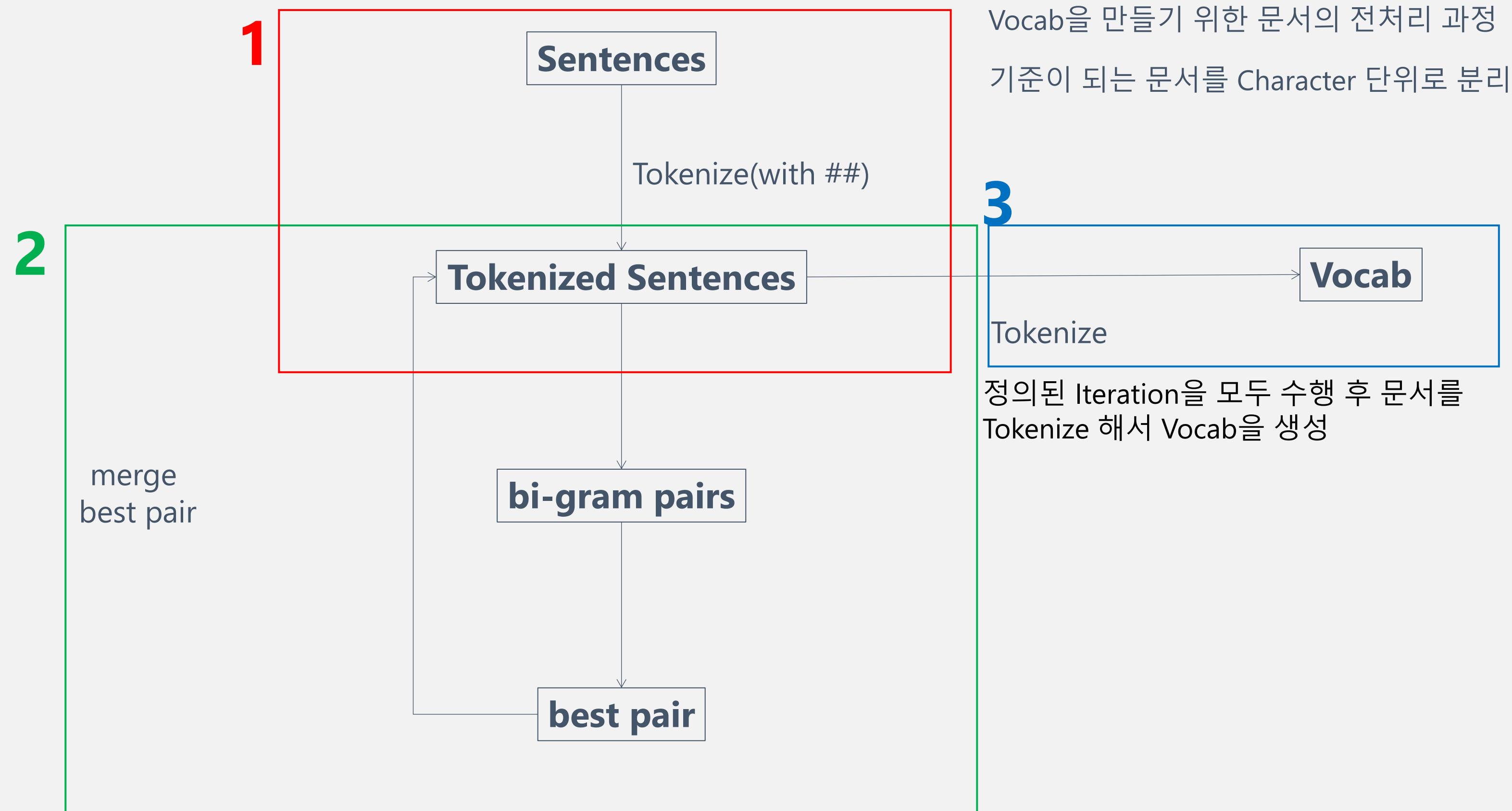
BPE vocabs

고양
##경찰
##서
##시
경찰
##관
##청
⋮

Methods – BERT의 WordPiece tokenizing

BERT

• BPE의 순서도



Vocab을 만들기 위한 문서의 전처리 과정
기준이 되는 문서를 Character 단위로 분리

정의된 Iteration을 모두 수행 후 문서를
Tokenize 해서 Vocab을 생성

Character 단위로 분리된 문서에서 가장 많이 등장하는 Bi-gram Pair를
찾고 합쳐주는 과정

Iteration을 정해 놓고 주어진 횟수 만큼 수행

Methods – BERT의 WordPiece tokenizing

BERT

- Tokenize

경찰청 철창살은 외철창살이고 검찰청 철창살은 쌍철창살이다

Sentence			Tokenized Sentence (iter=0)
경찰청	→경 찰 청	→	경 ##찰 ##청
철창살은	→철 창 살 은	→	철 ##창 ##살 ##은
외철창살이고	→외 철 창 살 이 고	→	외 ##철 ##창 ##살 ##이 ##고
검찰청	→검 찰 청	→	검 ##찰 ##청
철창살은	→철 창 살 은	→	철 ##창 ##살 ##은
쌍철창살이다	→쌍 철 창 살 이 다	→	쌍 ##철 ##창 ##살 ##이 ##다

- 같은 글자라고 맨 앞에 나오는 것과 아닌 것에는 차이가 있다고 가정
- BERT의 경우 뒷단어에 '##'을 붙여서 구별 e.g.) '철' ≠ '##철', '철창살' ≠ '##철창살'

Methods – BERT의 WordPiece tokenizing

BERT

- Building Vocab

Tokenized Sentence (iter=0)

경 ##찰 ##청
 철 ##창 ##살 ##은
 외 ##철 ##창 ##살 ##이 ##고
 검 ##찰 ##청
 철 ##창 ##살 ##은
 쌍 ##철 ##창 ##살 ##이 ##다

→

Vocab (iter=0) **Vocab 후보!**

경 ##찰 ##청
 철 ##창 ##살 ##은
 외 ##철 ##이 ##고
 검
 쌍 ##다

- Vocab의 생성은 정해진 Iteration을 모두 수행 후 Tokenized Sentence를 Tokenize해서 생성

Methods – BERT의 WordPiece tokenizing

BERT

- Bi-gram Pair Count

Vocab 후보 기준으로

Tokenized Sentence (iter=0)		Bi-gram pairs (iter=1)																								
경 ##찰 ##청	→	(경, ##찰), (##찰, ##청)																								
철 ##창 ##살 ##은	→	(철, ##창), (##창, ##살), (##살, ##은)																								
외 ##철 ##창 ##살 ##이 ##고	→	(외 ##철), (##철, ##창), (##창, ##살), (##살, ##이), (##이, ##고)																								
검 ##찰 ##청	→	(검, ##찰), (##찰, ##청)																								
철 ##창 ##살 ##은	→	(철, ##창), (##창, ##살), (##살, ##은),																								
쌍 ##철 ##창 ##살 ##이 ##다	→	(쌍 ##철), (##철, ##창), (##창, ##살), (##살, ##이), (##이, ##다)																								
<table><tr><td>(경, ##찰)</td><td>:1</td><td>(##살, ##은)</td><td>:2</td><td>(##이, ##고)</td><td>:1</td></tr><tr><td>(##찰, ##청)</td><td>:2</td><td>(외 ##철)</td><td>:1</td><td>(검, ##찰)</td><td>:1</td></tr><tr><td>(철, ##창)</td><td>:2</td><td>(##철, ##창)</td><td>:2</td><td>(쌍 ##철)</td><td>:1</td></tr><tr><td>(##창, ##살)</td><td>:4</td><td>(##살, ##이)</td><td>:2</td><td>(##이, ##다)</td><td>:1</td></tr></table>			(경, ##찰)	:1	(##살, ##은)	:2	(##이, ##고)	:1	(##찰, ##청)	:2	(외 ##철)	:1	(검, ##찰)	:1	(철, ##창)	:2	(##철, ##창)	:2	(쌍 ##철)	:1	(##창, ##살)	:4	(##살, ##이)	:2	(##이, ##다)	:1
(경, ##찰)	:1	(##살, ##은)	:2	(##이, ##고)	:1																					
(##찰, ##청)	:2	(외 ##철)	:1	(검, ##찰)	:1																					
(철, ##창)	:2	(##철, ##창)	:2	(쌍 ##철)	:1																					
(##창, ##살)	:4	(##살, ##이)	:2	(##이, ##다)	:1																					

Methods – BERT의 WordPiece tokenizing

BERT

- Merge Best pair

Best Pair: ##창 ##살 → **##창살**

Tokenized Sentence (iter=0)

경 ##찰 ##청

철 ##창 ##살 ##은

외 ##철 ##창 ##살 ##이 ##고

검 ##찰 ##청

철 ##창 ##살 ##은

쌍 ##철 ##창 ##살 ##이 ##다

Tokenized Sentence (iter=1)

경 ##찰 ##청

철 **##창살** ##은

외 ##철 **##창살** ##이 ##고

검 ##찰 ##청

철 **##창살** ##은

쌍 ##철 **##창살** ##이 ##다

Methods – BERT의 WordPiece tokenizing

BERT

•Building Vocab

Tokenized Sentence (iter=1)		Vocab 후보 업데이트	Vocab (iter=1)
경 ##찰 ##청		경	##찰 ##청
철 ##창살 ##은		철	##창살 ##은
외 ##철 ##창살 ##이 ##고	→	외	##철 ##이 ##고
검 ##찰 ##청		검	
철 ##창살 ##은			
쌍 ##철 ##창살 ##이 ##다		쌍	##다

- Vocab의 생성은 정해진 Iteration을 모두 수행 후 Tokenized Sentence를 Tokenize해서 생성

Methods – BERT의 WordPiece tokenizing

BERT

•Bi-gram Pair Count

Vocab 후보 기준으로

Tokenized Sentence (iter=1)		Bi-gram pairs (iter=2)
경 ##찰 ##청	→	(경, ##찰), (##찰, ##청)
철 ##창살 ##은	→	(철, ##창살), (##창살, ##은)
외 ##철 ##창살 ##이 ##고	→	(외 ##철), (##철, ##창살), (##창살, ##이), (##이, ##고)
검 ##찰 ##청	→	(검, ##찰), (##찰, ##청)
철 ##창살 ##은	→	(철, ##창살), (##창살, ##은),
쌍 ##철 ##창살 ##이 ##다	→	(쌍 ##철), (##철, ##창살), (##창살, ##이), (##이, ##다)
(경, ##찰)	:1	(외 ##철) :1 (검, ##찰) :1
(##찰, ##청)	:2	(##철, ##창살) :2 (쌍 ##철) :1
(철, ##창살)	:2	(##창살, ##이) :2 (##이, ##다) :1
(##창살, ##은)	:2	(##이, ##고) :1

Methods – BERT의 WordPiece tokenizing

BERT

- Merge Best pair

Best Pair: ##찰 ##청 → **##찰청**

Tokenized Sentence (iter=1)

경 ##찰 ##청

철 ##창살 ##은

외 ##철 ##창살 ##이 ##고

검 ##찰 ##청

철 ##창살 ##은

쌍 ##철 ##창살 ##이 ##다

Tokenized Sentence (iter=2)

→ 경 **##찰청**

→ 철 ##창살 ##은

→ 외 ##철 ##창살 ##이 ##고

→ 검 **##찰청**

→ 철 ##창살 ##은

→ 쌍 ##철 ##창살 ##이 ##다

- Best pair가 여러 개여도 하나만 선택

Methods – BERT의 WordPiece tokenizing

BERT

•Building Vocab

Tokenized Sentence (iter=2)		Vocab 후보 업데이트	Vocab (iter=2)
경 ##찰청		경	##찰청
철 ##창살 ##은		철	##창살 ##은
외 ##철 ##창살 ##이 ##고		외	##철 ##이 ##고
검 ##찰청	→	검	
철 ##창살 ##은			
쌍 ##철 ##창살 ##이 ##다		쌍	##다

- Vocab의 생성은 정해진 Iteration을 모두 수행 후 Tokenized Sentence를 Tokenize해서 생성

Methods – BERT의 WordPiece tokenizing

BERT

•Bi-gram Pair Count

Vocab 후보 기준으로

Tokenized Sentence (iter=2)		Bi-gram pairs (iter=3)
경 ##찰청	→	(경, ##찰청)
철 ##창살 ##은	→	(철, ##창살), (##창살, ##은)
외 ##철 ##창살 ##이 ##고	→	(외 ##철), (##철, ##창살), (##창살, ##이), (##이, ##고)
검 ##찰청	→	(검, ##찰청)
철 ##창살 ##은	→	(철, ##창살), (##창살, ##은),
쌍 ##철 ##창살 ##이 ##다	→	(쌍 ##철), (##철, ##창살), (##창살, ##이), (##이, ##다)

(경, ##찰청)	:1	(##철, ##창살)	:2	(쌍 ##철)	:1
(철, ##창살)	:2	(##창살, ##이)	:2	(##이, ##다)	:1
(##창살, ##은)	:2	(##이, ##고)	:1		
(외 ##철)	:1	(검, ##찰청)	:1		

Methods – BERT의 WordPiece tokenizing

BERT

- Merge Best pair

Best Pair: 철 ##창살 → **철창살**

Tokenized Sentence (iter=2)

경 ##찰 ##청

철 ##창살 ##은

외 ##철 ##창살 ##이 ##고

검 ##찰 ##청

철 ##창살 ##은

쌍 ##철 ##창살 ##이 ##다

Tokenized Sentence (iter=3)

경 ##찰청

→ **철창살** ##은

→ 외 ##철 ##창살 ##이 ##고

→ 검 ##찰청

→ **철창살** ##은

→ 쌍 ##철 ##창살 ##이 ##다

- 3 글자 이상의 패턴도 Iteration이 진행되면서 나타날 수 있음

Methods – BERT의 WordPiece tokenizing

BERT

•Building Vocab

Tokenized Sentence (iter=3)

경 ##찰청

철창살 ##은

외 ##철 ##창살 ##이 ##고

검 ##찰청

철창살 ##은

쌍 ##철 ##창살 ##이 ##다

→

Vocab (iter=3) **Vocab에 저장!**

경 ##찰청

철창살 ##은

외 ##철 ##창살 ##이 ##고

검

쌍 ##다

- Vocab의 생성은 정해진 Iteration을 모두 수행 후 Tokenized Sentence를 Tokenize해서 생성(매 iteration 마다 Vocab을 생성하는 것이 아님)

Methods – BERT의 WordPiece tokenizing

BERT

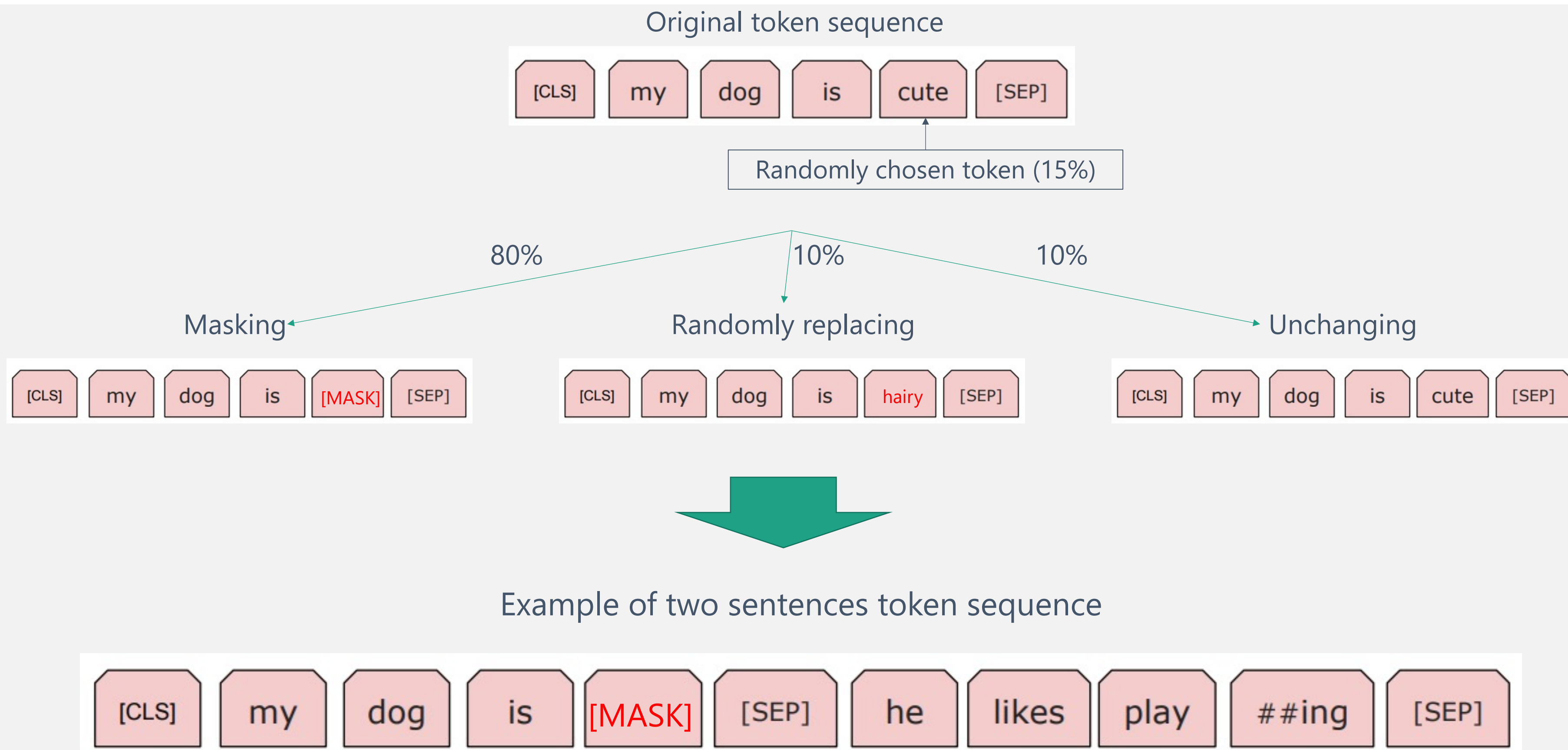
- BERT tokenization 예시
 - BPE로 subword로 분리했음에도 불구하고 vocab에 존재하지 않는 단어는 [UNK] 토큰으로 변환

```
INFO:tensorflow:tokens: [CLS] 유스티니아 ##뉴스의 분노 ##는 쯔 ##를 듯 ##했으나 수도 ##를 떠나는 실수를 한 바람에 필리 ##피 ##>
쿠스가 콘스탄티노폴리스 ##를 차지하였고 유스티니아 ##뉴스는 11월 4일 붙잡혀 처형 ##당했다 ##. [SEP] ##여 ##버 ##렸고 유스티니아
##뉴스의 아들 6 [MASK] 티 ##베리우스 ##도 블라 ##게 ##르나이 성당에서 처형 ##당 ##했는데 이로써 [MASK] ##클리우스 왕조의 혈 ##
통 ##은 완전히 끊 ##어졌다 ##. [UNK] 전동차의 공기 ##제동 방식 가운데 하나이다 ##. 1868년 ##에 조지 웨 ##스팅 [MASK] 발명 ##하>
였다 ##. 순수 ##하게 공기 ##압 제어로 동작 ##하는 것은 크게 단행 ##운전용인 [UNK] 연결 ##운전용인 비상 ##변 부착 [UNK] 나뉜다 #
#. 개요 SM 공급 [UNK] Reser ##v ##o ##ir ##: [UNK] 불리는 가 ##압 ##된 공기 탱 ##크로부터 ##, 운전 ##대 ##까지 연결된 공기 ##저
장관으로 불리는 공기관 ##을 통해 공기 ##압 ##을 공급 ##하여 제동 ##변을 조작 ##해 개폐 ##하는 것으로 ##, [UNK] Air P ##ip ##e #
#: [UNK] 불리는 브레이크 실린 ##더 직결 공기관 ##에 가 ##압 ##하여 제동력을 얻는 매우 단순한 제동 시스템이다 ##. 웨 ##스팅 ##하
우스가 불인 공식 이름은 [UNK] a ##ir b ##ra ##ke ##/ ##M ##otor c ##ar ##: 전동차 641 [UNK] 세계 여러 나라의 단행 ##운전차량에
널리 보급 [MASK] 노 ##면 ##전차 ##의 경우 아직도 영업 일 ##선에서 많이 쓰이고 있다. 다만 이 시스템은 구조가 간단 ##하며 동작 ##
이 신 ##속 ##하고 확실 ##하지만 ##, 공기관 ##이 파 ##손 ##될 경우 제동 ##이 걸 ##리지 않게 되는 위험 ##이 있기 때문에 ##, 보안
##상 연결 운전에는 사용할 수 없다는 단 ##점이 있다. SME SM [MASK] ##에서 문제가 된 열차 ##분리 [MASK] 발생 등의 대응 ##책 ##으>
로 비상용 자동 ##공기 ##제동 [UNK] 그 지 ##령에 이용하는 [UNK] [UNK] 함께 설치한 [UNK] a ##ir b ##ra ##ke ##/ ##M ##otor c ##ar
##/ ##E ##mer ##gen [MASK] v ##al ##ve ##: 전동차 ##용 비상 ##변 부착 직통 ##공기 ##제동 ##. 모터 ##가 없는 트 ##레일 ##러 ##>
용 ##은 ST ##E 혹은 [UNK] 웨 ##스팅 ##하우스 ##사가 개발 ##하여 2~ ##3 ##량 정도의 단편성 ##용으로 보급 ##되었다 ##. 이 SME ##>
는 원형 ##인 SM ##과 같은 직통제동기구를 이루고 있지만 ##, [MASK] 저장기에 [MASK] 공기 ##탱 ##크가 원 ##공기 [UNK] Reser [MASK]
##o ##ir ##: [UNK] 불리며 공기 저장관 ##도 원 ##공기 [UNK] [MASK] ##v ##o ##ir P ##ip ##e ##: [UNK] 불리고 있다. 이것은 SME [M
ASK] [MASK] ##제동부에 비상 ##제동의 동력 ##원을 공급 ##하는 보조 공기 ##저장기가 존재 ##하기 때문에 ##, [MASK] 구별 ##하기 위
##함이다 ##. 비상 ##변 ##에는 평 ##상 ##시는 490 ##k ##P ##a ##의 압 ##력이 가해 ##지고 있어 ##, 긴급 ##할 때 ##나 비상 ##변 ##
의 호 ##스가 파 ##열 ##되었을 때도 비상 ##제동 ##이 작동 ##한다 ##. 비상 ##제동 ##은 자동 ##공기 ##제동 ##과 같이 보조 공기 ##>
저장기의 공기를 배출시키 ##는 것으로 작동시키기 때문에 ##, 안전성이 항상 ##되었다 ##. 브레이크의 가 ##감 ##압 ##은 [MASK] ##래>
의 SM 방식 ##과 달리 가 ##감 ##압 ##의 속도가 언제나 정해져 있다. 제동 단계는 [UNK] [UNK] [UNK] [UNK] [UNK] 압 ##력을 [UNK] [UN
K] 압 ##력을 [UNK] 이렇게 4 ##가지 ##가 있다. 제2차 십자군 ( ##114 ##7년 - [UNK] 제1차 십자군 [MASK] 이후 팔레스타 ##인의 십자>
```


Methods – Masking 기법

BERT

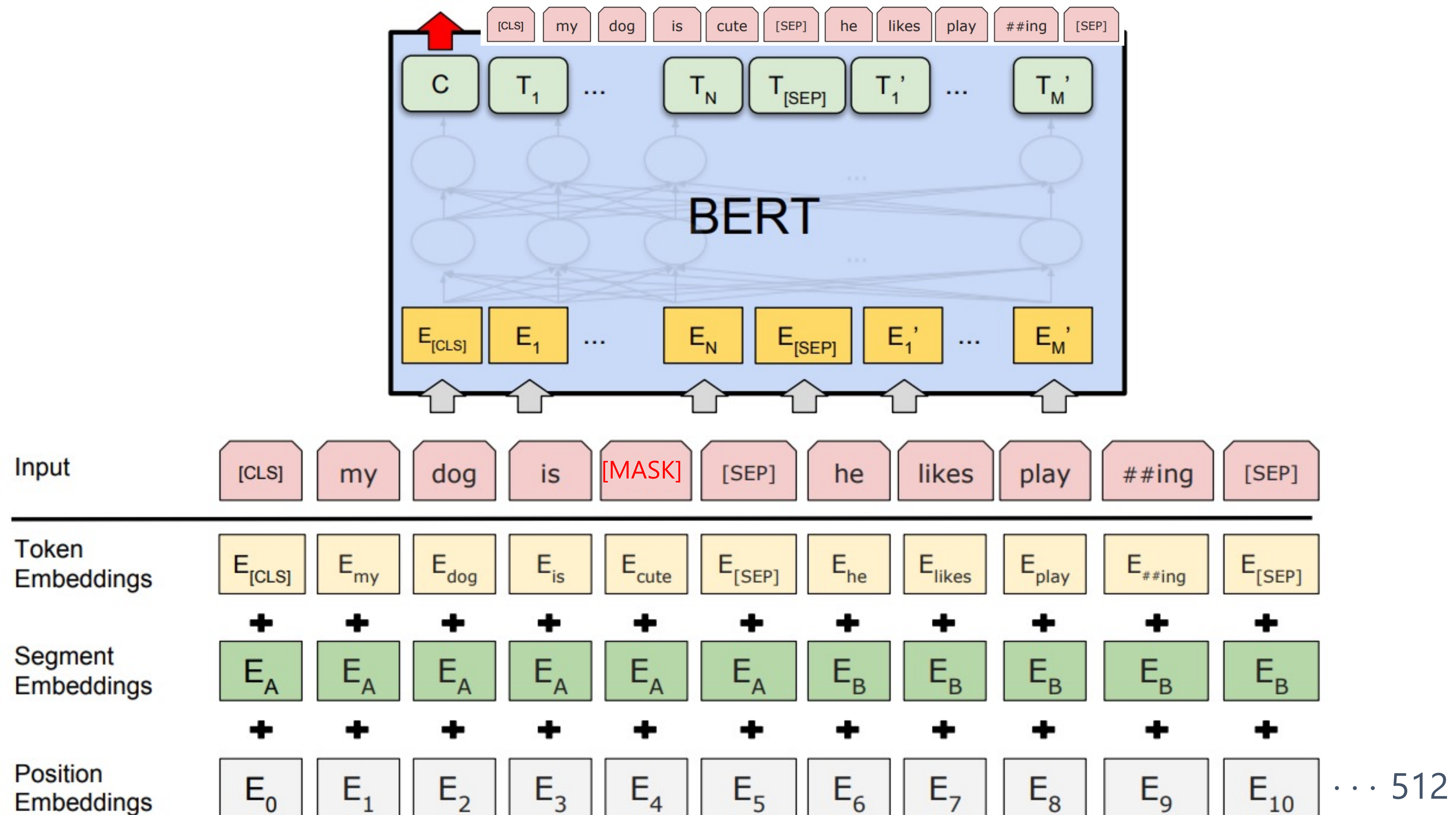
- 데이터의 tokenizing
 - Masked language model (MLM): input token을 일정 확률로 masking



Methods – Masking 기법

BERT

Class Label = **IsNext** or **NotNext**

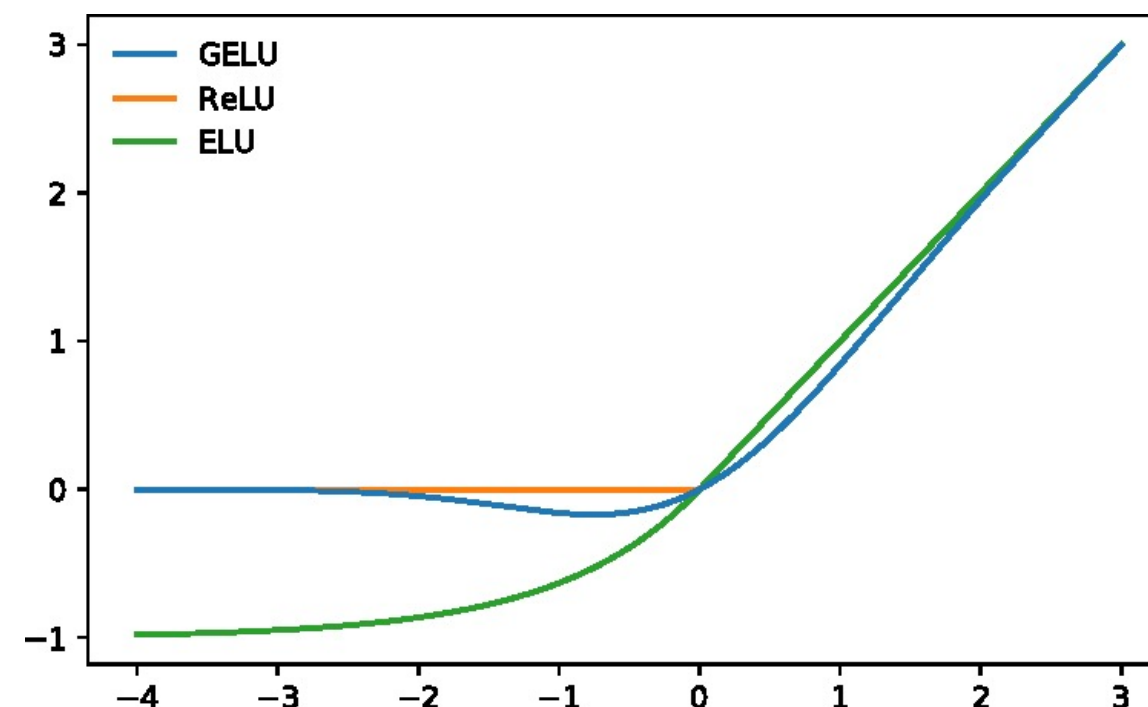


Methods

BERT

• Training options

- Train batch size: 256 sequences (256 sequences * 512 tokens = 128,000 tokens/batch)
- Steps: 1M
- Epoch: 40 epochs
- Adam learning rate: $1e-4$
- Weight decay: 0.01
- Drop out probability: 0.1
- Activation function: GELU



• Environmental setup

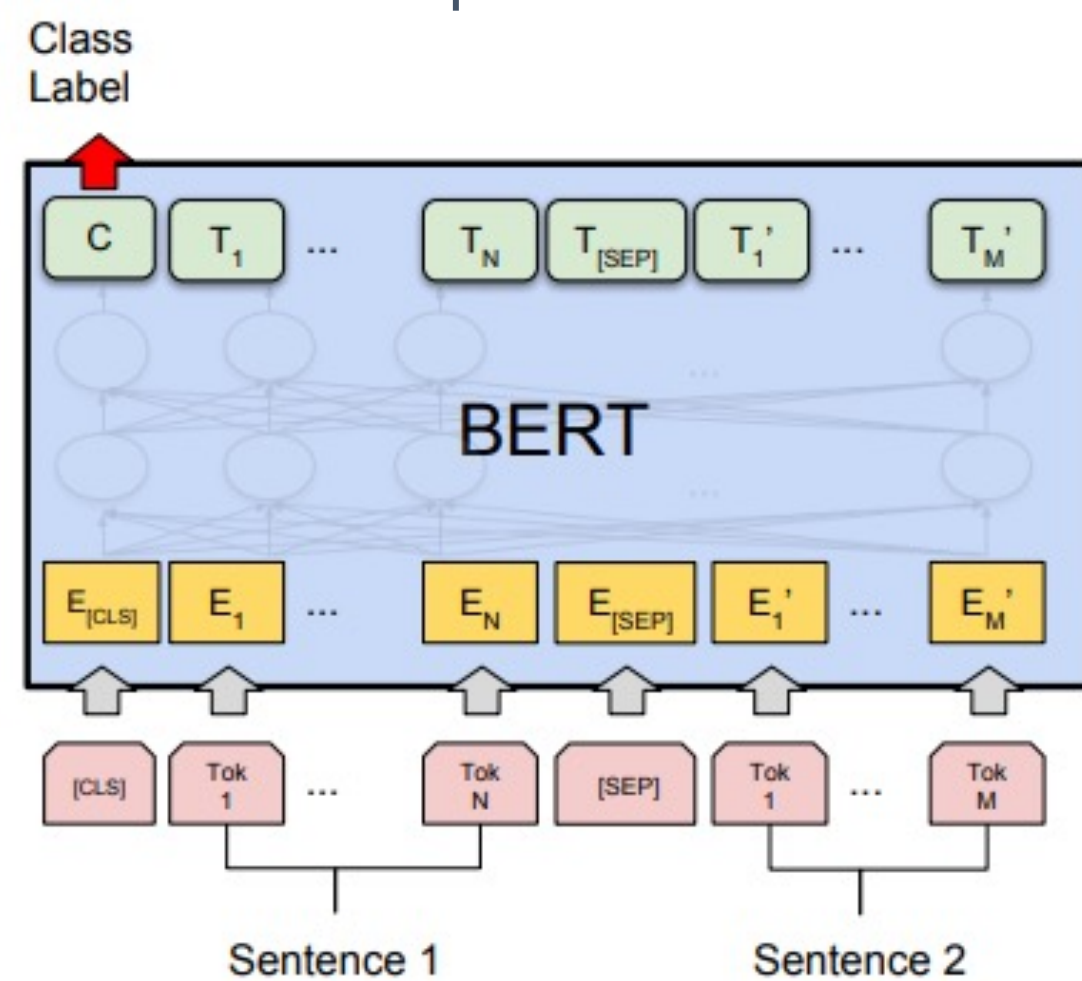
- BERT_{BASE}: 4 Cloud TPUs (16 TPU chips total)
- BERT_{LARGE}: 16 Cloud TPUs (64 TPU chips total) \approx 72 P100 GPU
- Training time: 4 days

Methods

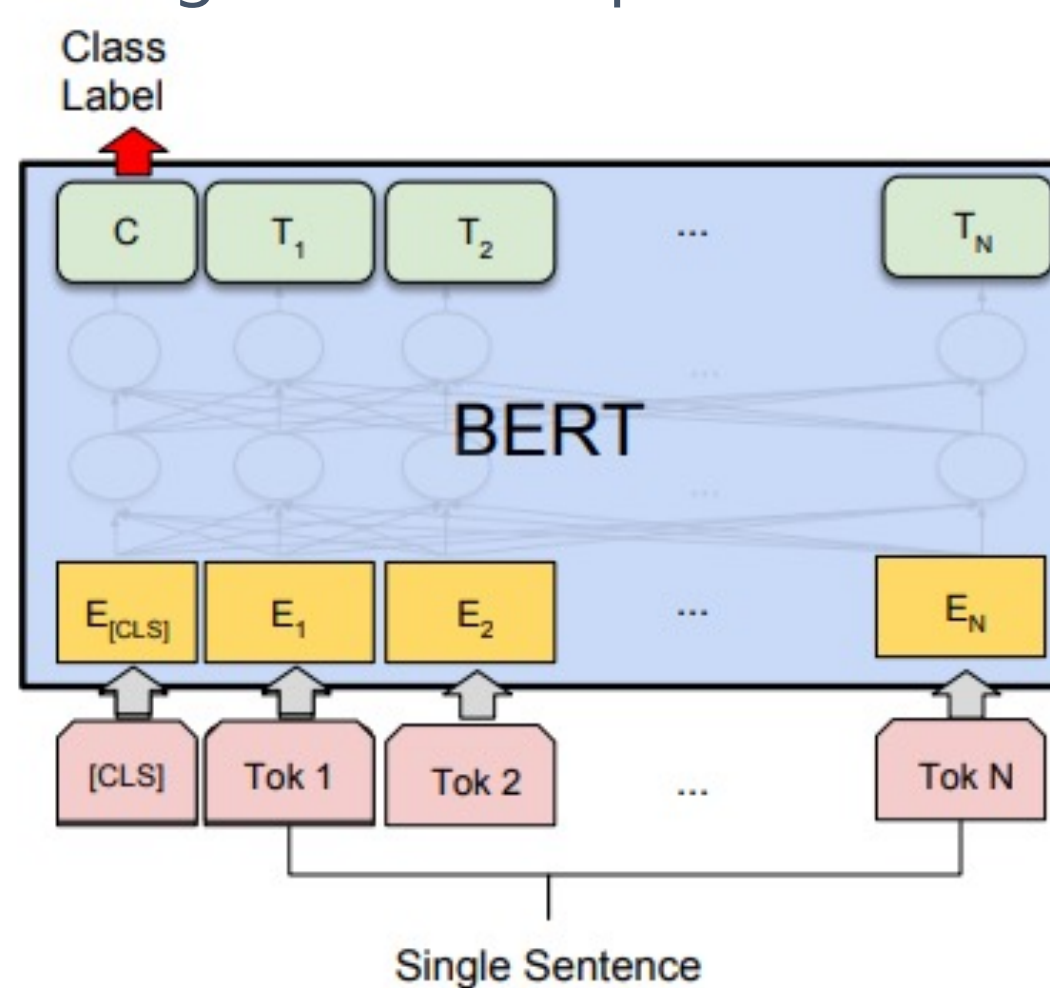
BERT

- NLP 실험

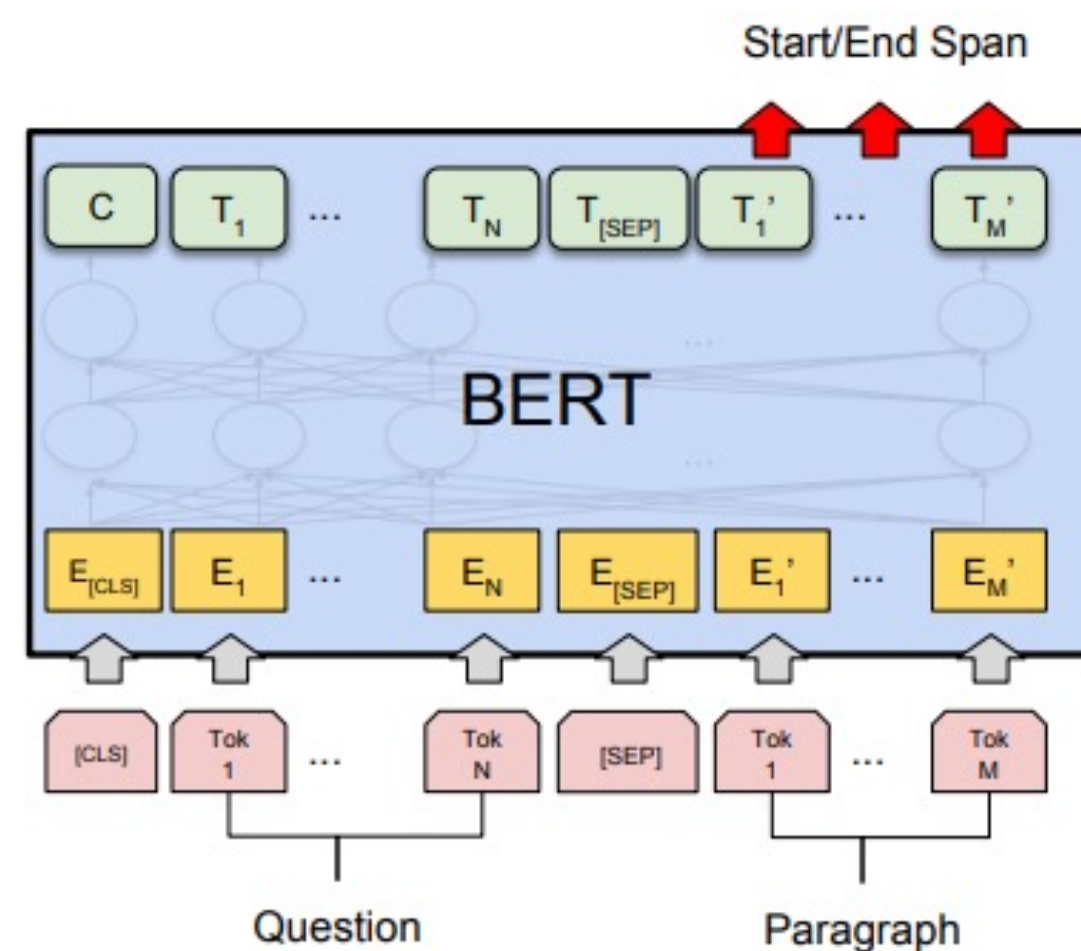
Sentence pair classification



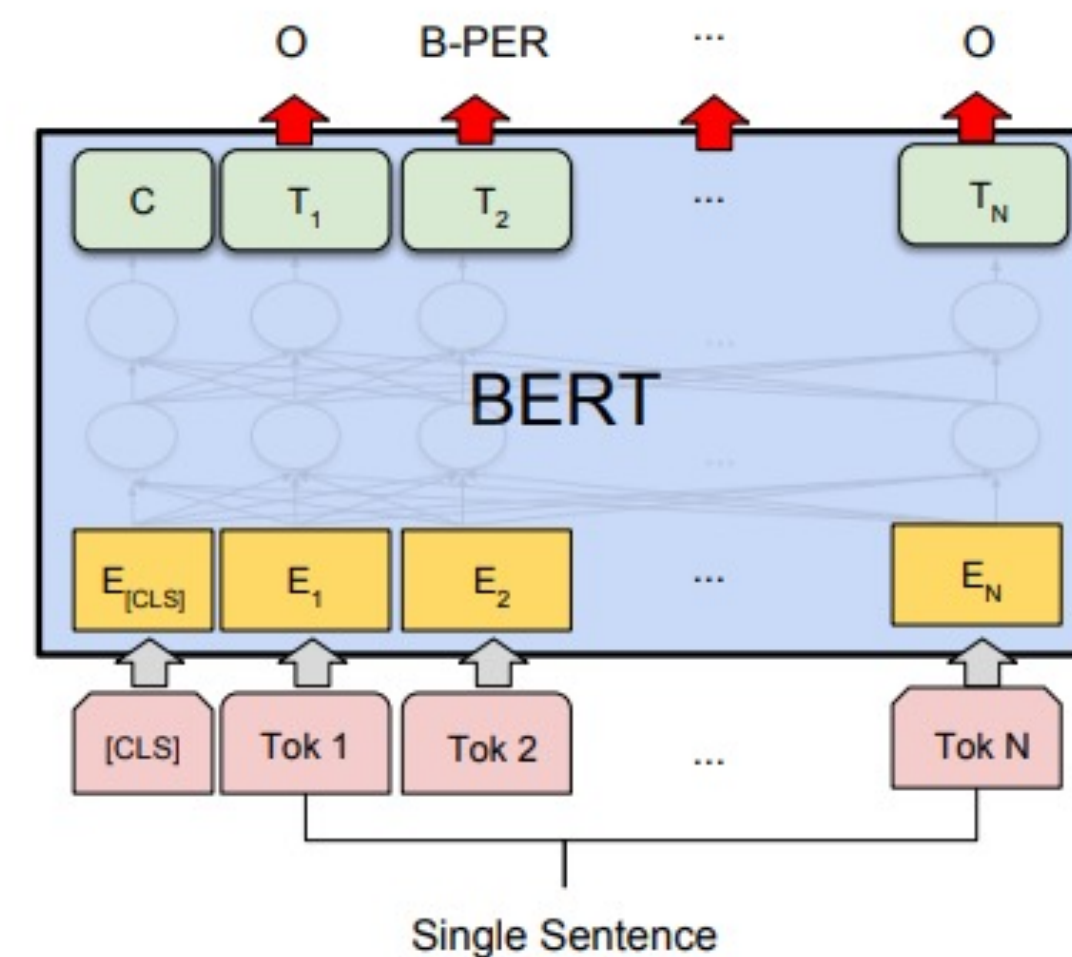
Single sentence pair classification



Question and answering (SQuAD v1.1)



Single sentence tagging



Results

BERT

- GLUE test results

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
BERT _{BASE}	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
BERT _{LARGE}	86.7/85.9	72.1	91.1	94.9	60.5	86.5	89.3	70.1	81.9

- SQuAD v1.1

System	Dev		Test	
	EM	F1	EM	F1
Leaderboard (Oct 8th, 2018)				
Human	-	-	82.3	91.2
#1 Ensemble - nlnet	-	-	86.0	91.7
#2 Ensemble - QANet	-	-	84.5	90.5
#1 Single - nlnet	-	-	83.5	90.1
#2 Single - QANet	-	-	82.5	89.3
Published				
BiDAF+ELMo (Single)	-	85.8	-	-
R.M. Reader (Single)	78.9	86.3	79.5	86.6
R.M. Reader (Ensemble)	81.2	87.9	82.3	88.5
Ours				
BERT _{BASE} (Single)	80.8	88.5	-	-
BERT _{LARGE} (Single)	84.1	90.9	-	-
BERT _{LARGE} (Ensemble)	85.8	91.8	-	-
BERT _{LARGE} (Sgl.+TriviaQA)	84.2	91.1	85.1	91.8
BERT _{LARGE} (Ens.+TriviaQA)	86.2	92.2	87.4	93.2

- Input Question:

Where do water droplets collide with ice
crystals to form precipitation?

- Input Paragraph:

... Precipitation forms as smaller droplets
coalesce via collision with other rain drops
or ice crystals within a cloud. ...

- Output Answer:

within a cloud

Results

BERT

- Named Entity Recognition (CoNLL-2003)

System	Dev F1	Test F1
ELMo+BiLSTM+CRF	95.7	92.2
CVT+Multi (Clark et al., 2018)	-	92.6
BERT _{BASE}	96.4	92.4
BERT _{LARGE}	96.6	92.8

- SWAG

System	Dev	Test
ESIM+GloVe	51.9	52.7
ESIM+ELMo	59.1	59.2
BERT _{BASE}	81.6	-
BERT _{LARGE}	86.6	86.3
Human (expert) [†]	-	85.0
Human (5 annotations) [†]	-	88.0

A girl is going across a set of monkey bars. She

- (i) jumps up across the monkey bars.
- (ii) struggles onto the bars to grab her head.
- (iii) gets to the end and stands on a wooden plank.
- (iv) jumps up and does a back flip.

BERT 적용 실험 - 감성 분석

BERT

- 네이버 영화 리뷰 코퍼스 (<https://github.com/e9t/nsmc>) 로 감성 분석 진행
- 학습: 150,000 문장 / 평가: 50,000 문장 (긍정: 1, 부정: 0)

label	sentence
0	뭐야 이 평점들은.... 나쁘진 않지만 10점 짜리는 더더욱 아니잖아
0	지루하지는 않은데 완전 막장임... 돈주고 보기에...
0	3D만 아니어도 별 다섯 개 줬을텐데.. 왜 3D로 나와서 제 심기를 불편하게 하죠??
1	음악이 주가 된, 최고의 음악영화
0	진정한 쓰레기
0	마치 미국애니에서 튀어나온듯한 창의력없는 로봇디자인부터가,고개를 젓게한다

모델	Accuracy
KoNLPy + DenseNet (Keras)	85.3
Kakao BERT-Kor + khaiii	88.8
BERT Multi-lingual pretrained model	88.3

BERT 적용 실험 – QnA 문장 유사도 (이진 분류 기반)

29

BERT

- 디지털 동반자 패러프레이징 질의 문장 데이터를 이용하여 질문-질문 데이터 생성 및 학습
- 학습: 3,401 문장 쌍 (유사 X: 1,700개, 유사 O: 1,701개)
- 평가: 1,001 문장 쌍 (유사 X: 500개, 유사 O: 501개)

label	sentence 1	sentence 2
0	포화지방산이 많은 음식은?	아토피 피부염이 있는 아기를 씻길 때 좋은 비누는 무엇인가요?
1	임신 29주 태아의 신장은 어떻습니까?	임신 29주일 때 태아의 크기는?
0	산모가 술을 마셨을 경우 모유 수유는 언제 하는 것이 좋은가요?	임산부가 입덧을 할 때 비타민B6을 왜 먹나요?

모델	Accuracy
BERT Multi-lingual pretrained model	98.3

BERT 적용 실험 – QnA 문장 유사도 (이진 분류 기반)

30

BERT

- Saltlux ADAM 일반 질의 데이터를 이용해 IRQA 모사 데이터 구축
- 500 문장에 대한 BERT semantic search (<https://github.com/hanxiao/bert-as-service>)

중고차 대출시 필요서류 알고 싶습니다.



[0.2132, 0.1544, 0.8163] 768차원

⋮

500 문장 테스트

전세자금대출을 받았는데, 추가로 대출을 더 받을수 있는지 궁금합니다.

겨울쿨톤입니다. 겨울쿨 립스틱이 찾기어려워요 겨울 톤 가진 분들
뭐 쓰시나요?

여름에 조개같은거 먹고 질병에 안 걸리려면 어떻게 해야될까요?

중고차 대출을 받으려고요. 필요한 서류가 뭔가요?

⋮

1만 문장과 비교

모델	Top1 Accuracy
ADAM IRQA algorithm	80.0
Sent2Vec (https://arxiv.org/abs/1703.02507)	82.7
BERT Multi-lingual pretrained model (이진분류)	58.4

BERT 적용 실험 – QnA 문장 유사도 (문장 벡터 기반)

BERT

- Sent2Vec 이용, 500개 입력 문장과 가장 유사한 10개의 문장 후보 추출 (top 10개 이내 98% 존재)
- 10개 문장과 입력 문장을 BERT로 유사 유/무 비교 (이진 분류)

입력 문장	유사 문장	Cos similarity
무협소설 추천좀 해주세요	판타지소설 추천좀 해주세요	0.80263
	고전문학 소설 추천해주세요	0.658433
	현대문학 소설 추천해주세요	0.645056
	대표적인 무협소설 뭐 있을까요?	0.634835
	처음읽는데도 거부감이없고 흥미로운 무협소설 추천바랍니다	0.632528
	요새 재미있는 무협소설 뭐 있나요?	0.631053
	PC게임 추천좀 해주세요	0.61363
	고전문학 소설 읽을만한 소설 소개좀 해주세요	0.594516
	현대문학 소설 읽을만한 소설 소개좀 해주세요	0.59369
	패션잡지 추천 좀 해주세요	0.573995

BERT 적용 실험 – QnA 문장 유사도 (문장 벡터 기반)

32

BERT

- ADAM 일반 질의 데이터 중, most similar Top 10 문장들을 이용해 학습 데이터 구축
- 학습: 22,766 문장 / 평가: 5,000 문장

label	sentence 1	sentence 2
1	중고차 대출시 필요서류 알고 싶습니다.	중고차 대출을 받으려고요. 필요한 서류가 뭔가요?
0	전세를 살고 있는데 전세대출을 받았거든요. 다른 곳으로 이사가면 돈을 상환해야 합니까?	전세자금대출을 받았는데, 추가로 대출을 더 받을수 있는지 궁금합니다.
1	겨울쿨톤 이 무난하게 잘 쓸수 있는 립스틱 어떻게 있을까요?	겨울쿨톤입니다. 겨울쿨 립스틱이 찾기어려워요 겨울 톤 가진 분들 뭐 쓰시나요?
1	비브리오 패혈증을 예방하는 방법이 뭔가요?	여름에 조개같은거 먹고 질병에 안 걸리려면 어떻게 해야될까요?
0	디지털 디톡스가 뭐죠?	세상에서 가장 비싼그림이 뭐죠?

•
•
•

모델	Top1 Accuracy
ADAM IRQA algorithm	80.0
Sent2Vec (https://arxiv.org/abs/1703.02507)	82.7
BERT Multi-lingual pretrained model	58.4
Sent2Vec + BERT Multi-lingual pretrained model	90.0

BERT 적용 실험 – 관계 추출 실험

BERT

- KAIST가 구축한 Silver data 사용 (1명의 전문가가 annotation)
- 학습: 985,806 문장 / 평가: 100,001 문장
- 총 81개 label (관계 없음 포함)

Sbj	label	Obj	sentence
미겔 데 세르반테스	majorWork	돈 키호테	1570년에 기사도 로망스 소설 돈 키호테를 쓴 스페인 작가 미겔 데 세르반테스가 나폴리에서 스페인 병사로 복무했었다.
겨울연가	NA	배용준	대한민국의 드라마인 겨울연가에서 배용준의 차량으로 등장해 인지도를 얻었으나, 실제로는 판매가 부진하였다.
라타쿱가	locatedIn	에콰도르	코토팍시 국제공항은 에콰도르 코토팍시 주 라타쿱가에 있는 국제공항이다.

•
•
•

모델	Accuracy
PCNN + Attention + RL (2018 SOTA)	69.5
BERT Multi-lingual pretrained model	85.7

BERT 적용 실험 – 개체명 인식

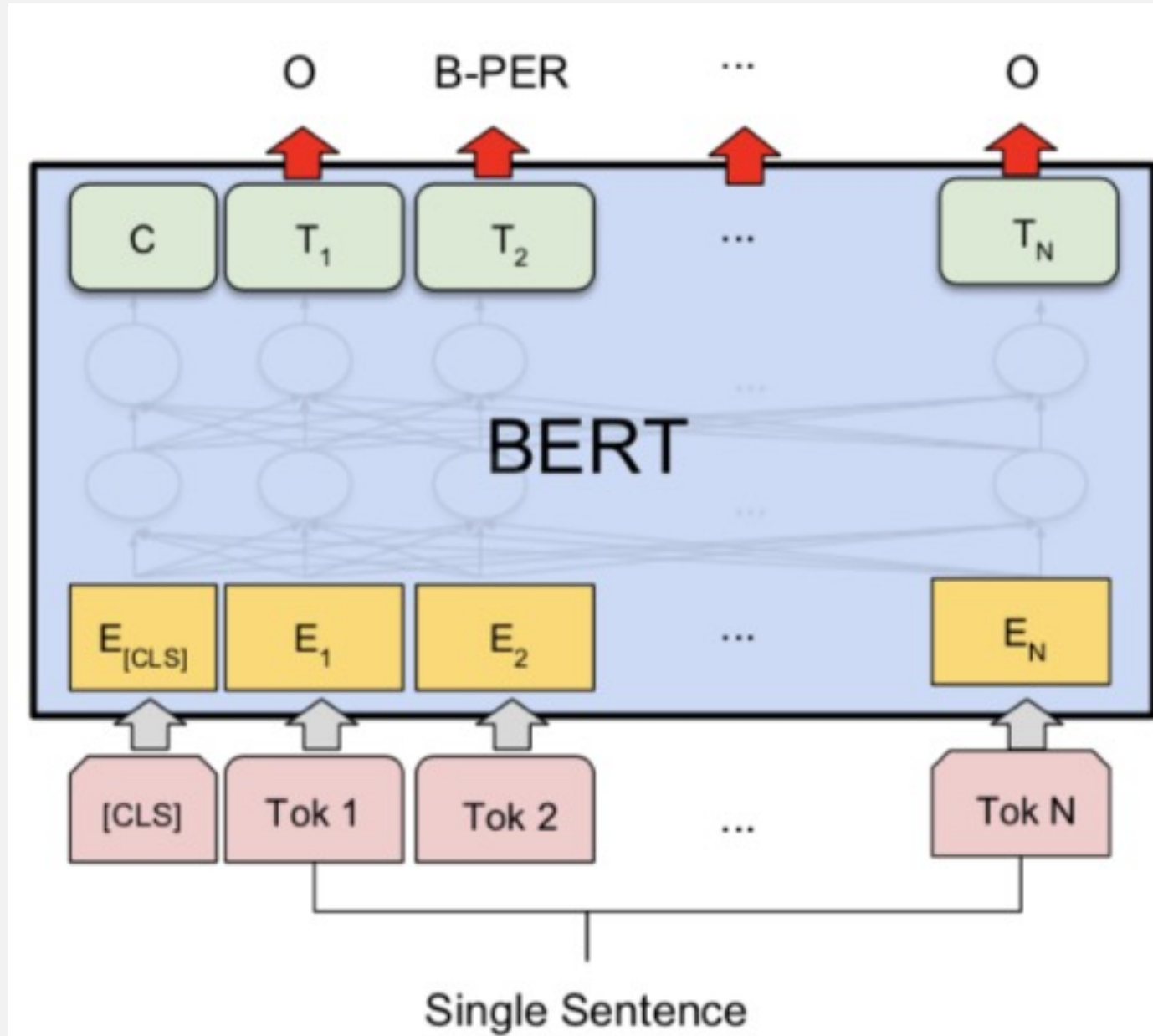
BERT

- 버트 모델에서 맨 마지막에 sequence tagging을 위한 layer를 쌓아서 구현
- 공개된 다국어 모델 활용, 음절 단위로 입력을 하도록 구현 (<https://github.com/kyzhouhzau/BERT-NER>)

<증권거래소:OG>에 주식을 상장하였다



증	##권	##거	##래	##소	##에	주	##식	##을	상	##장	##하	##였	##다
B-OG	I-OG	I-OG	I-OG	I-OG	O	O	O	O	O	O	O	O	O



BERT 적용 실험 – 개체명 인식

BERT

- Saltlux ADAM 개체명 인식 데이터를 활용하여 학습 및 평가 진행 (정보통신단체표준 TTA.KO-10.0852)
- 학습: 95,787 문장 / 평가: 10,503 문장

기존 개체명 인식기 성능

	개체명	Precision	Recall	F1 score
Artifacts	AF	88.87	76.05	81.96
Animal	AM	92.76	78.21	84.82
Civilization	CV	88.1	71.07	78.67
Date	DT	93.8	92.95	93.38
Event	EV	67.53	60.09	63.59
Study_field	FD	84.6	73.73	78.79
Location	LC	95.71	92.15	93.9
Material	MT	87.04	81.07	83.95
Organization	OG	79.37	85.17	82.17
Person	PS	86.32	83.15	84.7
Plant	PT	95.86	82.47	88.66
Quantity	QT	91.1	93.24	92.16
Time	TI	81.02	79.28	80.14
Term	TM	76.23	74.71	75.46
Theory	TR	72.17	52.91	61.06

버트 기반 모델 성능

	개체명	Precision	Recall	F1 score
	AF	87.12	88.13	87.61
	AM	85.14	86.2	85.67
	CV	79.77	88.79	84.04
	DT	96.32	96.54	96.43
	EV	86.64	75.12	80.46
	FD	76.29	83.9	79.91
	LC	94.49	95.66	95.07
	MT	83.12	91.07	86.91
	OG	89.63	87.37	88.48
	PS	89.42	91.8	90.59
	PT	83.29	92.4	87.61
	QT	95.74	94.09	94.91
	TI	91.3	83.61	87.29
	TM	82.45	84.23	83.33
	TR	79.65	76.18	77.88

BERT 적용 실험 – 기계독해 실험

BERT

- LG CNS가 공개한 한국어 QA 데이터 셋, KorQuAD (<https://korquad.github.io/>)

규모: TRAIN 60,407 / DEV 5,774 / TEST 3,898

주소: <https://korquad.github.io/>



Context

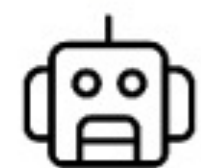
세계 AI 경진대회, LG CNS "4위"



LG CNS가 세계적 인공지능(AI) 학회인
인공신경망학회(NeurlPS) 주최 AI경진
대회에서 톱5에 진입하는 성과를 거뒀다
고 20일 밝혔다. LG CNS는 NeurlPS의
AI경진대회 중 '이미지 인식 AI 대회'에
서 미국 카네기멜론대(1위), 중국 칭화대
(2위), 캐나다 몬트리올 고등기술대(3위)
에 이어 4위에 올랐다. 톱5 수상팀 가운
데 기업은 LG CNS가 유일하다.
...<후략>

LG CNS 뉴스 일부, 한국경제

Reading



Machine

Reading

Question

세계적인 AI 경진대회에서
톱5에 든 기업 어디야?



User

Comprehension

Answer

"LG CNS" 입니다.

BERT 적용 실험 – 기계독해 실험

BERT

- LG CNS가 공개한 한국어 QA 데이터 셋, KorQuAD (<https://korquad.github.io/>)
- Wikipedia article에 대해 10,645 건의 문단과 66,181 개의 질의응답 쌍으로, Training set 60,407 개, Dev set 5,774 개의 질의응답쌍으로 구분
- Google 공개 버전 BERT 다국어 모델 KorQuad 실험



모델	EM	F1
Human performance	80.17	91.20
Saltlux ADAM MRC	71.88	83.00
BERT Multi-lingual pretrained model (KorQuAD, 어절 단위 tokenizing)	70.19	89.85
BERT Multi-lingual pretrained model (KorQuAD, 음절 단위 tokenizing)	81.78	90.75

BERT 적용 실험 – TOEIC 문제 풀이

BERT

🔗 TOEIC-BERT

76% Correct rate with ONLY Pre-Trained BERT model in TOEIC!!

This is project as topic: TOEIC(Test of English for International Communication) problem solving using pytorch-pretrained-BERT model. The reason why I used huggingface's [pytorch-pretrained-BERT model](#) is for pre-training or to do fine-tune more easily. I've solved the only blank problem, not the whole problem. There are two types of blank issues:

1. Selecting Correct Grammar Type.

Q) The music teacher had me _ scales several times.

1. play (Answer)
2. to play
3. played
4. playing

2. Selecting Correct Vocabulary Type.

Q) The wet weather _ her from going playing tennis.

1. interrupted
2. obstructed
3. impeded
4. discouraged (Answer)

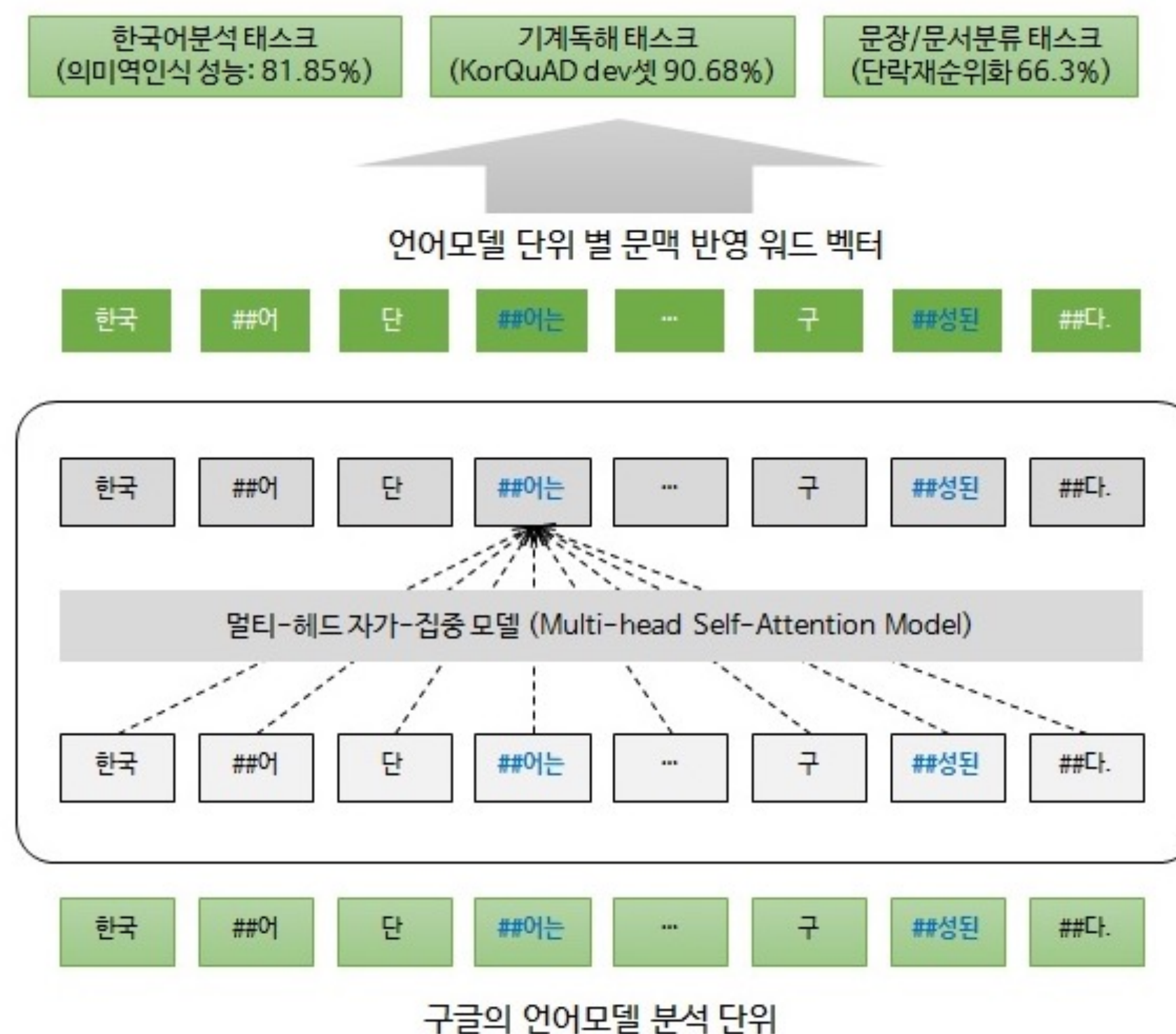
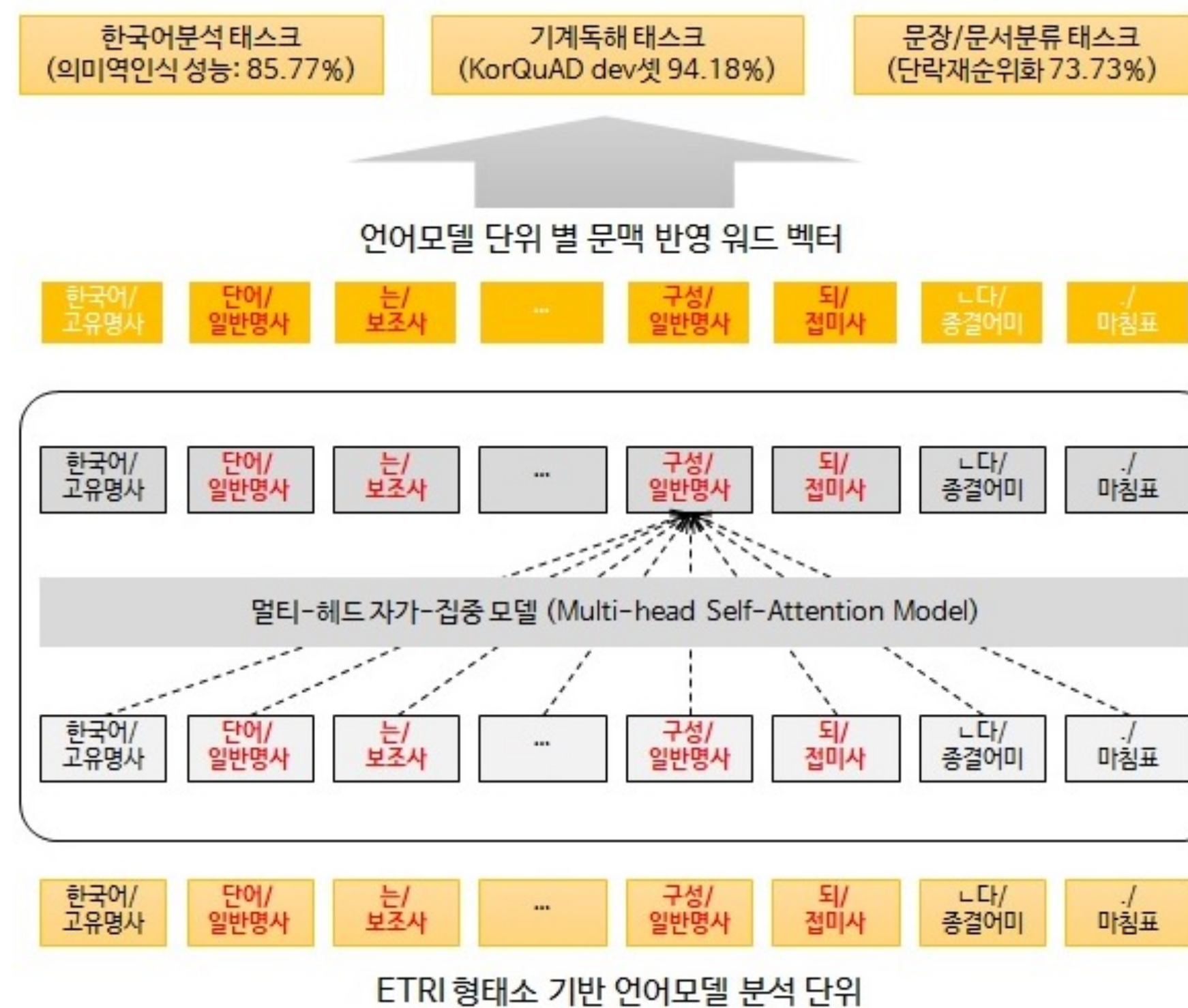
https://github.com/graycode/toeicbert?fbclid=IwAR2hoCQE02CaR00m-RZCHwQM_kYd1LgxxMSrucYSTtA52ZUhtvq5i_G2tFk

한국어 BERT

ETRI에서 공개한 KorBERT의 특징과, 한국어로 BERT를 학습할 때 중요한 점을 살펴봅니다

ETRI KorBERT

한국어 BERT



예문: 한국어 단어는 형태소로 구성된다.

〈ETRI 형태소 기반 언어모델과 구글 언어모델 비교〉

ETRI KorBERT

한국어 BERT

배포 모델	세부 모델	세부 내용	모델 파라미터
KorBERT	Korean_BERT_Morphology	<ul style="list-style-type: none"> 학습데이터: 23GB 원시 말뭉치 (47억개 형태소) 형태소분석기: 본 OpenAPI 언어분석 중, 형태소분석 API 딥러닝 라이브러리: pytorch, tensorflow 소스코드: tokenizer 및 기계독해(MRC), 문서분류 예제 Latin alphabets: Cased^(대소문자 구분) 	30349 vocabs, 12 layer, 768 hidden, 12 heads,
	Korean_BERT_WordPiece	<ul style="list-style-type: none"> 학습데이터: 23GB 원시 말뭉치 딥러닝 라이브러리: pytorch, tensorflow 소스코드: tokenizer Latin alphabets: Cased^(대소문자 구분) 	30797 vocabs, 12 layer, 768 hidden, 12 heads,

```

./SF_ 20688850
다 /EF_ 17194907
하 /XSV_ 15918761
ㄴ /ETM_ 13942053
을 /JKO_ 13654182
있 /EP_ 12912273
의 /JKG_ 12012373
에 /JKB_ 11360511
이 /VCP_ 8986000
이 /JKS_ 8938636
, /SP_ 8614178
는 /JX_ 8594108
를 /JKO_ 8310256
어 /EC_ 7491203
은 /JX_ 7238132

```

```

INFO:tensorflow:tokens: [CLS] 우리 /NP_ 은 하 /NNG_ 중 심 /NNG_ 의 /JKG_ 궁 수 /NNG_ >
자 리 /NNG_ A/SL_ */SW_ 불 려 흘 /NNP_ 의 /JKG_ 직 접 /MAG_ 관 측 /NNG_ 을 /JKO_ 시 /NNG_
도 /JX_ 하 /VV_ 고 /EC_ 있 /VX_ 는 /ETM_ 것 /NNB_ 은 /JX_ ?/SF_ [SEP] 불 려 흘 /NNG_ >
은 /JX_ 그 /MM_ 성 질 /NNG_ 상 /XSN_ 이 톤 /NNG_ 상 /XSN_ 의 /JKG_ 호 칭 /NNG_ 복 사 /NNG_
를 /JKO_ 제 외 /NNG_ 하 /XSV_ 면 /EC_ 직 접 /NNG_ 적 /XSN_ 이 /VCP_ ㄴ /ETM_ 전 자 기 /NNG_
복 사 /NNG_ 를 /JKO_ 방 출 /NNG_ 하 /XSV_ 지 /EC_ 양 /VX_ 으 며 /EC_ , /SP_ 때 문 /NNB_ >
에 /JKB_ 불 려 흘 /NNG_ 을 /JKO_ 찾 /VV_ 는 /ETM_ 전 체 /NNG_ 둘 리 학 자 /NNG_ 를 /XSN_ >
은 /JX_ 간 접 /NNG_ 적 /XSN_ 관 측 /NNG_ 수 단 /NNG_ 에 /JKB_ 의 존 /NNG_ 하 /XSV_ 아 야 /EC_
하 /VX_ 하 /VX_ ㄴ 다 /EF_ ./SF_ 예 컨 대 /MAG_ 불 려 흘 /NNG_ 주 위 /NNG_ 의 /JKG_ 중 력
/NNG_ 적 /XSN_ 상 호 /NNG_ 작 용 /NNG_ 을 /JKO_ 통 하 /VV_ 아 /EC_ 불 려 흘 /NNG_ 의 /JKG_
존 재 /NNG_ 를 /JKO_ 알 아 내 /VV_ 는 /ETM_ 것 /NNB_ 이 /JKS_ 그 러 /X R_ 하 /XSA_ ㄴ /ETM_
간 접 /NNG_ 적 /XSN_ 관 측 /NNG_ 수 단 /NNG_ 중 /NNB_ 하 나 /NR_ 이 /VCP_ 다 /EF_ ./SF_ 그
러 나 /MAJ_ M IT/SL_ 헤 이 스 렉 /NNP_ 전 문 대 /NNG_ 의 /JKG_ 사 건 /NNG_ 의 /JKG_ 지 평
선 /NNG_ 망 원 경 /NNG_ (/SS_ E H T/SL_ )/SS_ 는 /JX_ 우리 /NNG_ 은 하 /NNG_ 중 심 /NN
G_ 의 /JKG_ 궁 수 /NNG_ 자 리 /NNG_ A/SL_ */SW_ 불 려 흘 /NNG_ 에 /JKB_ 대 하 /VV_ ㄴ /ET
M_ 직 접 /MAG_ 관 측 /NNG_ 을 /JKO_ 시 도 /NNG_ 중 /NNB_ 이 /VCP_ 며 /EC_ , /SP_ 2016/SN_
년 /NNB_ 초 /NNB_ 에 /JKB_ 는 /JX_ 최 초 /NNG_ 로 /JKB_ 사 건 /NNG_ 의 /JKG_ 지 평 선 /NNG
을 /JKO_ 포 착 /NNG_ 하 /XSV_ ㄴ /ETM_ 사 진 /NNG_ 을 /JKO_ 얻 /VV_ 을 /ETM_ 수 /NNB_ 있
/VV_ 을 /ETM_ 것 /NNB_ 으 로 /JKB_ 예 상 /NNG_ 하 /XSV_ 고 /EC_ 있 /VX_ 다 /EF_ ./SF_ 궁
수 /NNG_ 자 리 /NNG_ A/SL_ */SW_ 의 /JKG_ 사 건 /NNG_ 의 /JKG_ 지 평 선 /NNG_ 바 로 /MAG
바 짚 /NNG_ 에 /JKB_ 자 기 /NP_ 장 이 /XSN_ 존 재 /NNG_ 하 /XSV_ = /ETM_ 것 /NNB_ 이 /VC
P_ 라는 /ETM_ 이 톤 /NNG_ 적 /XSN_ 예 상 /NNG_ 은 /JX_ 2015/SN_ 년 /NNB_ E H T/SL_ 에 /J
KB_ 의 하 /VV_ 아 /EC_ 사 실 /NNG_ 로 /JKB_ 밝 히 /VV_ 어 /EC_ 지 /VX_ 었 /EP_ 다 /EF_ ./SF
[SEP]

```


ETRI KorBERT를 위한 Open 형분석기

한국어 BERT

- ETRI KorBERT의 입력은 형분석 이후 데이터가 입력되어야 함

대분류	소분류	세분류			
(1) 체언	명사(NN)	일반명사(NNG)			목적격조사(JKO)
		고유명사(NNP)			부사격조사(JKB)
		의존명사(NNB)			호격조사(JKV)
	대명사(NP)	대명사(NP)			인용격조사(JKQ)
		수사(NR)			수사(NR)
(2) 용언	동사(VV)	동사(VV)	(6) 의존형태	어미(EM)	보조사(JX)
	형용사(VA)	형용사(VA)			접속조사(JC)
	보조용언(VX)	보조용언(VX)			선어말어미(EP)
	지정사(VC)	긍정지정사(VCP)			종결어미(EF)
		부정지정사(VCN)			연결어미(EC)
(3) 수식언	관형사(MM)	성상 관형사(MMA)		접미사(XS)	명사형전성어미(ETN)
		지시 관형사(MMD)			관형형전성어미(ETM)
		수 관형사(MMN)			체언접두사(XPN)
	부사(MA)	일반부사(MAG)			명사파생접미사(XSN)
		접속부사(MAJ)			동사파생접미사(XSV)
(4) 독립언	감탄사(IC)	감탄사(IC)	(7) 기호	일반기호(ST)	형용사파생접미사(XSA)
(5) 관계언	격조사(JK)	주격조사(JKS)			어근(XR)
		보격조사(JKC)			마침표, 물음표, 느낌표(SF)
		관형격조사(JKG)			쉼표, 가운뎃점, 콜론, 빗금(SP)
					따옴표, 괄호표, 줄표(SS)
		줄임표(SE)			
		불임표(물결)(SO)			
		기타 기호(SW)			
		외국어(SL)		외국어(SL)	
		한자(SH)		한자(SH)	
		숫자(SN)		숫자(SN)	
		분석불능범주(NA)		분석불능범주(NA)	

ETRI KorBERT를 위한 Open 형분석기

한국어 BERT

- ETRI KorBERT의 입력은 형분석 이후 데이터가 입력되어야 함
- ETRI 형분석기의 경우, 일 5,000건으로 제한
- Saltlux ADAM 형분석기와 카카오 Khaii 형분석기의 경우, ETRI 형분석기 태그 셋과 동일
(품사 태그 세트 표준(TTA.KO-11.0010), 세종21계획의 결과물인 '형태분석 말뭉치 구축 지침(Ver. 2005-1))

아버지가방에들어가신다.



형분석기	형태소 분석 결과
ETRI	아버지가방/NNG + 예/JKB + 들어가/VV + 시/EP + ㄴ 다/EF + ./SF
Khaii	아버지/NNG + 가/JKS + 방/NNG + 예/JKB + 들어가/VV + 시/EP + ㄴ 다/EF + ./SF
ADAM	아버지/NNG + 가/JKS + 방/NNG + 예/JKB + 들어가/VV + 시/EP + ㄴ 다/EP + ./SF 아버지/NNG + 가/JKS + 방/NNG + 예/JKB + 들어가/VV + 신/EP + 다/EF + ./SF
Mecab	아버지/NNG + 가/JKS + 방/NNG + 예/JKB + 들어가/VV + 신다/EP+EC
Twitter	아버지/Noun + 가방/Noun + 예/Josa + 들어가신/Verb + 다/Eomi

한국어 BERT 학습

한국어 BERT

- 대용량 코퍼스 확보
 - Dump wiki + news 색인 데이터 (30GB)

모델	어절 수
Google 영어 BERT model	약 3.3억 어절
한국어 BERT 실험	약 3.4억 어절

- 실험 장비
 - V100 GPU
 - 32GB
 - 112TFLOPS
 - 예상 소요시간: 약 8주
 - Google 단일 TPU
 - 64GB
 - 180 TFLOPS
 - V100 1대 속도의 약 8배
 - 시간당 \$4.50
 - 예상 소요시간: 약 2주

한국어 BERT 학습

한국어 BERT

- 한국어 BERT baseline model 학습
 - 1/32 학습 소요 시간 (V100 GPU)



KorQuAD 실험결과

모델	EM	F1
Human performance	80.17	91.20
Saltlux ADAM MRC	71.88	83.00
BERT Multi-lingual pretrained model	81.78	90.75
한국어 BERT (1/32 학습)	66.06	79.79



성능이 안좋다!

한국어 BERT 학습

한국어 BERT

- BERT 성능에 영향을 미치는 요인
 - Corpus 사이즈
 - Corpus 도메인
 - Corpus tokenizing (어절, BPE, 형태소)
 - Vocab 사이즈 (영어 model: 30,522 vocabs, 다국어 model: 119,547 vocabs)

모델	예제
어절 + BPE (22만 vocab)	멕시코 ##부터 브라질 ##까지 그리고 카리브 해 섬 지역에서 발견된다 ##. 8 ##종으로 이루어져 있다.
어절 + BPE (4만 vocab)	날 ##씬 ##한 박쥐 ##류로 특이하게 ##도 긴 꼬리와 이름이 함 ##축하 ## 는 바와 같이 깔 ##때 ##기 모양의 귀를 갖고 있다.
형태소 분리	멕시코 부터 브라질 까지 그리고 카리브 해 섬 지역 에서 발견 된 다 . 8 종 으로 이루어져 있 다 .
형태소 분리 + 형태소 tag	멕시코/NNP 부터/JX 브라질/NNP 까지/JX 그리고/MAJ 카리브/NNP 해 /NNG 섬/NNG 지역/NNG 에서/JKB 발견/NNG 된/XSV 다/EF ./SF 8/SN 종/NNG 으로/JKB 이루어져/VV 있/VX 다/EF ./SF
형태소 분리 + 원형 복원	멕시코 부터 브라질 까지 그리고 카리브 해 섬 지역 에서 발견 되 ㄴ 다 . 8 종 으로 이루어지 어 있 다 .
형태소 분리 + 동사 원형 복원	멕시코 부터 브라질 까지 그리고 카리브 해 섬 지역 에서 발견 되다 . 8 종 으로 이루다 지다 있다.

한국어 BERT 학습

한국어 BERT

- BERT 학습
 - 코퍼스: 6.3GB, 약 6.7억 어절
 - 배치 사이즈: 128
 - 학습 steps: $\approx 400,000$
- Vocab size: $\approx 30,000$
- KorQuAD로 평가

모델	전처리	Tokenizing	Exact matching	F1
BERT _{base}	X	Word piece	67.23	86.39
BERT _{morp}	형태소 분리	형태소 단위	84.46	92.13
BERT _{morp+wp}	형태소 분리	Word piece	85.94	93.35
BERT _{morp+tag}	형태소 분리 +Tag	형태소 단위	81.81	91.86
BERT _{morp_r+tag+wp}	형태소 원형복원+ 형태소 분리 +Tag	Word piece	83.62	93.60

Advanced BERT Model

한국어 BERT

- KBQA에서 가장 중요한 entity 정보가 기존 BERT에서는 무시
- Entity linking을 통한 주요 entity 추출 및 entity tag 부착
- Entity embedding layer의 추가
- 형태소 분석을 통해 NNP와 entity 우선 chunking masking

원본 문장

이순신은 조선 중기의 무신이자 임진왜란 당시 조선 수군을 통솔했던 지휘관이자 구국영웅으로, 자는 여해(汝諧)이며, 시호는 충무공(忠武公)이다.



Entity 후보 추출

조선, 조선 중기, 조선 수군, 이순신, 임진왜란, 충무공



주요 entity 추출

이순신, 임진왜란, 조선, 조선 수군, 충무공



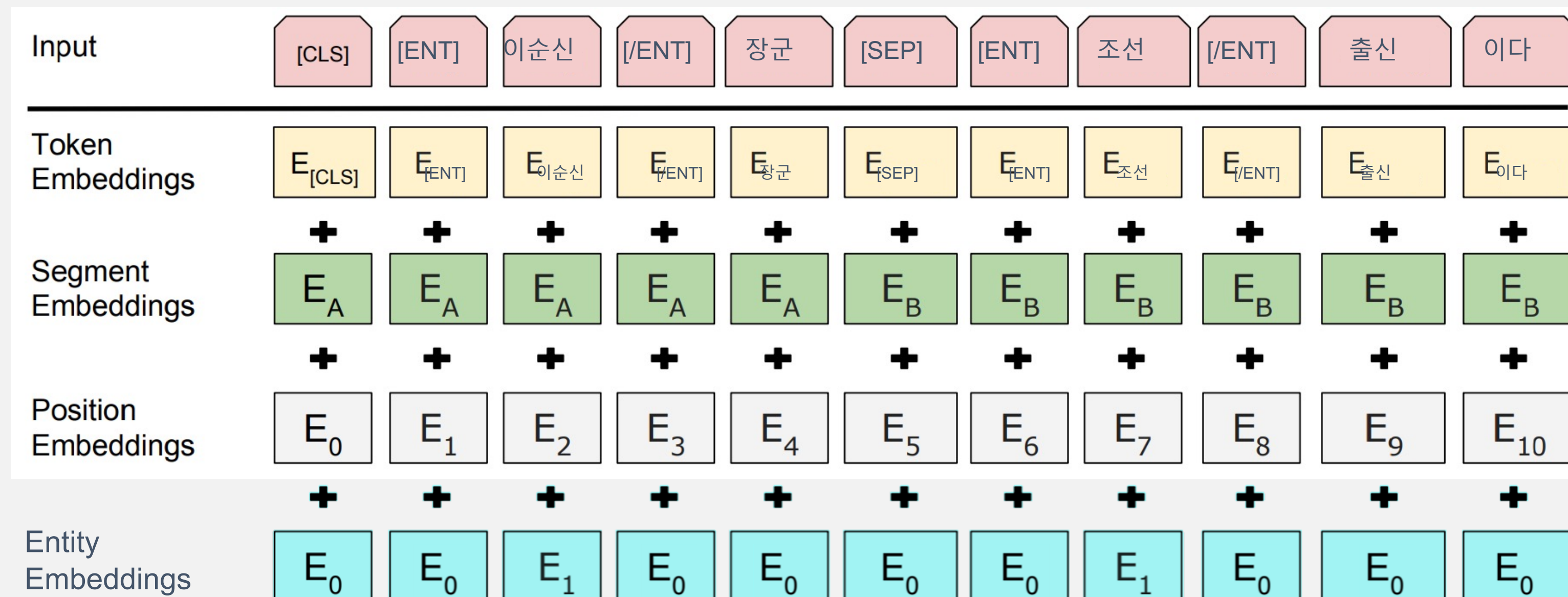
전처리

[ENT]이순신[/ENT]은 [ENT]조선[/ENT] 중기의 무신이자 [ENT]임진왜란[/ENT] 당시 [ENT]조선 수군[/ENT]을 통솔했던 지휘관이자 구국영웅으로, 자는 여해(汝諧)이며, 시호는 [ENT]충무공[/ENT](忠武公)이다.

Advanced BERT Model

한국어 BERT

- KBQA에서 가장 중요한 entity 정보가 기존 BERT에서는 무시
- Entity linking을 통한 주요 entity 추출 및 entity tag 부착
- Entity embedding layer의 추가
- 형태소 분석을 통해 NNP와 entity 우선 chunking masking



Advanced BERT Model

한국어 BERT

- 학습 데이터: 2019년 06월 20일 Wiki dump (약 4,700만 어절)
- Batch: 128
- Sequence length: 512
- Training steps: 300,000 (대략 10 epochs)

	KorQuAD dev set	
모델	Exact matching	F1
BERT _{base}	64.51	83.76
BERT _{Entity}	78.13	87.25



적은 학습 데이터로도 BERT의 성능 향상