

딥러닝을 활용한 영어문제 풀이

영어가 안되면 ~ 투빅스쿨.com

전종섭 이영걸 신용재
박성진 박경빈 김동환 이민주



INDEX

01



Introduction

02



Data

03



Models

04



Conclusion

01. Introduction

✓ 주제선정 배경 및 목적

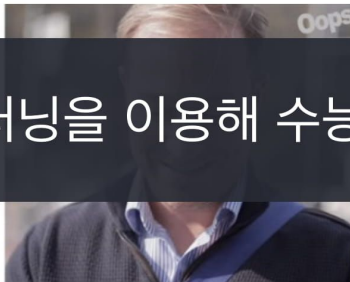


01. Introduction

✓ 주제선정 배경 및 목적

2019 수능 '최고난도' 영어 문제 풀다 '멘붕' 온 외국인들 (영상)

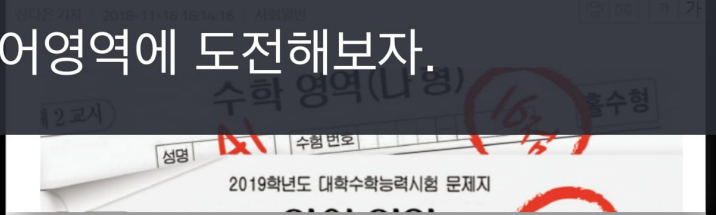
33. Heritage is concerned with the ways in which very selective material artefacts, mythologies, memories and traditions become treasures for the present. The contents, interpretations and representations of this resource are selected according to the demands of the present, an imagined future, or a desired past. It follows too that the meanings and memory and tradition are defined in the present. Heritage is more concerned with meanings than material artefacts. It is the former that give value, either cultural or financial, to the latter and explain why they have been selected from the near infinity of the past. In turn, they may later be discarded as the demands of present societies change, or even, as is presently occurring in the former Eastern Europe, when pasts have to be reinvented to reflect new presents. Thus



딥러닝을 이용해 수능 영어영역에 도전해보자.

AI '2019 수능' 풀었다...영어 12점·수학 16점

인공지능에 '2019 수능' 풀게해보니...
추론·사고력 문제 전혀 대응 못해



02. Data

✓ TOEIC

PART. 5

Q) The most recent economic crisis resulted in the near collapse of Smart Savings Bank, a highly _____ financial institution in Asia and Europe.

(A) recognizes
(B) recognized
(C) has recognized
(D) recognizing

PART. 6

Questions 120-123 refer to the following memo.

TO: All staff
FROM: Anthony Ricci
SUBJECT: New sale
DATE: July 8

You've no doubt heard that we ---120--- a sale on all our flat-screen televisions. It's been advertised heavily on the radio. I just want to give you a few pointers on how to sell these products. When customers enter our store, greet them ---121--- and then tell them about the sale. For those of you working in the television aisle, be sure to ---122--- conversations with browsing customers by saying that the sale only lasts until the end of the week. Once you've provided this information, you can tell people about the features of our various televisions. If they're interested, turn those televisions on so they can get a look at the picture quality for themselves. ---123---

120. (A) to hold (B) held (C) had held (D) are holding
121. (A) courteous (B) courteously (C) courteousness (D) courtesies
122. (A) withhold (B) interrupt (C) conclude (D) initiate
123. (A) This typically impresses customers. (B) The new one is sold out, unfortunately. (C) They know that this deal lasts for a limited time. (D) The lucky winner will be announced next Wednesday.

02. Data

✓ TOEIC

_____ to secure your export products, we recommend that you have them insured before shipment

Sentence	Label
<u>In order</u> to secure your export products, we recommend that you have them insured before shipment	True
<u>So that</u> to secure your export products, we recommend that you have them insured before shipment	False
<u>Though</u> to secure your export products, we recommend that you have them insured before shipment	False
<u>Because</u> to secure your export products, we recommend that you have them insured before shipment	False

02. Data

✓ CoLA (The Corpus of Linguistic Acceptability)

"I demand that John pay more, the more he **eats**."



Linguistically Acceptable

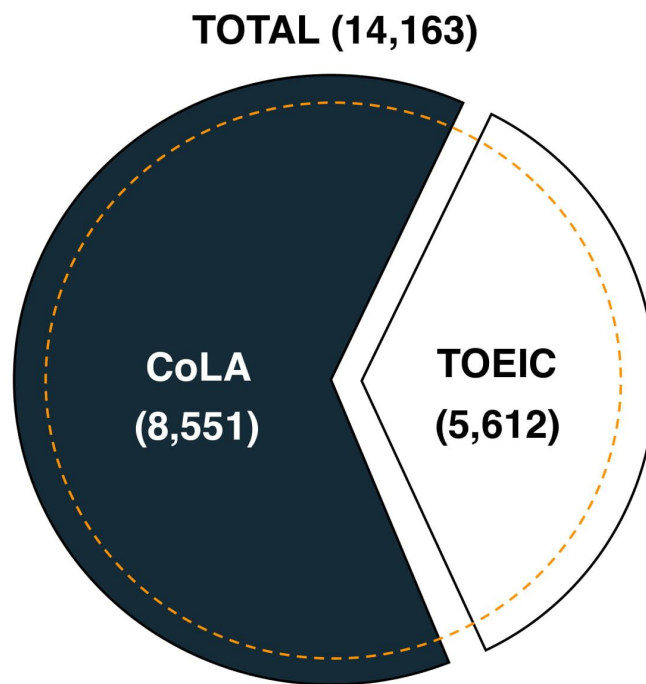
"I demand that John **pays** more, the more he **eat**."



Linguistically Unacceptable

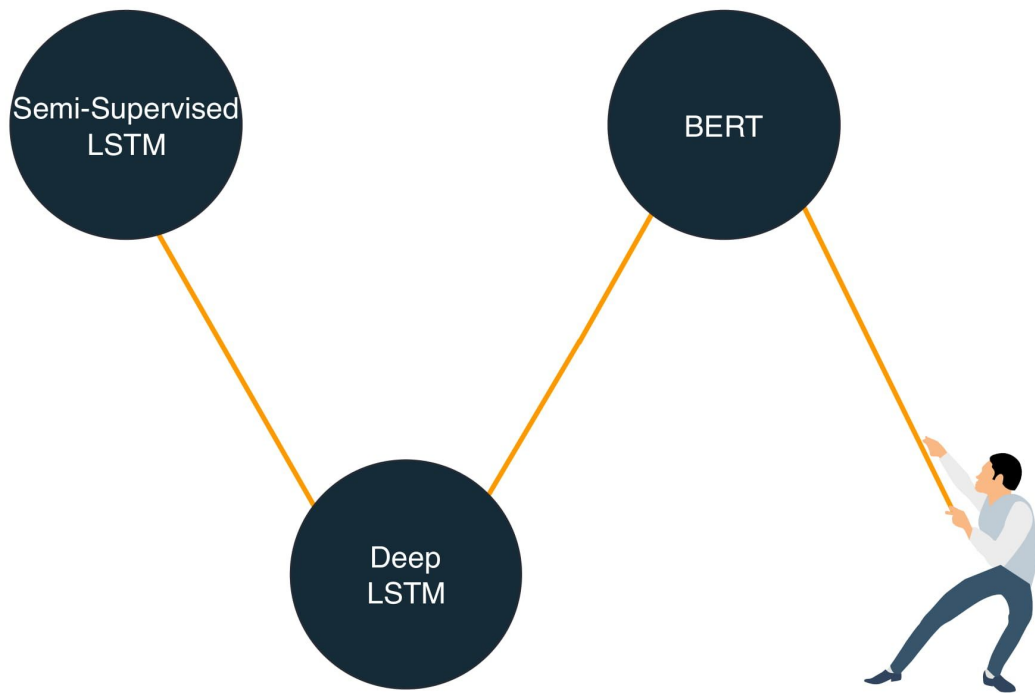
02. Data

✓ Total Dataset



03. Models

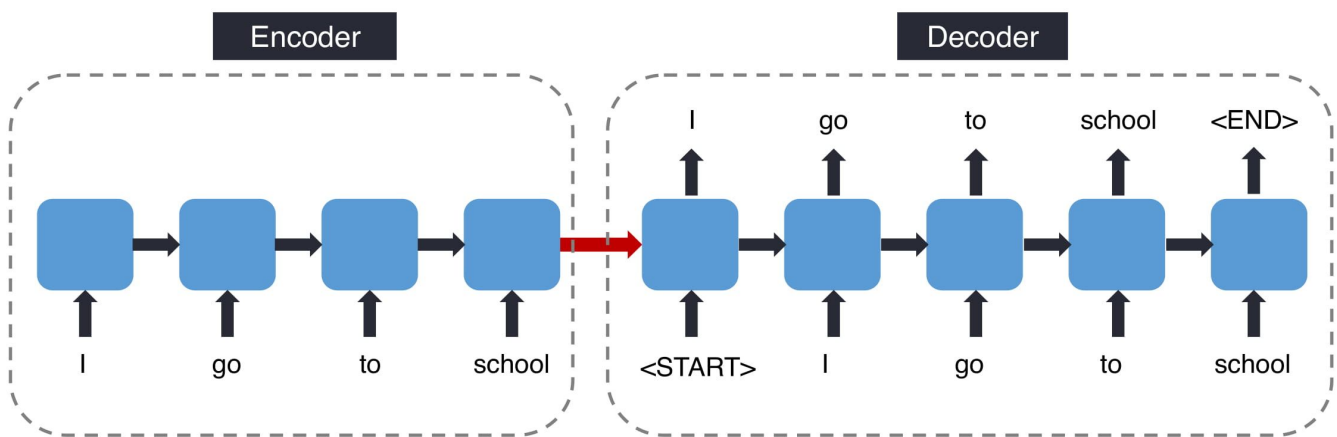
✓ Models



03. Models

✓ Model 1: LSTM

- Semi-supervised Sequence Learning (2015)

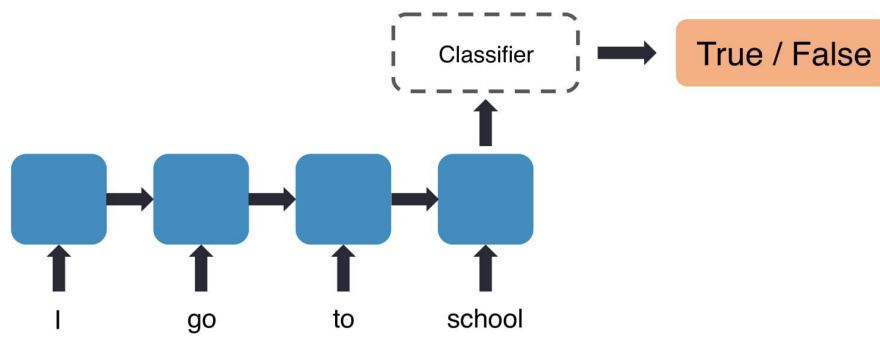


데이터의 sentence를 사용해 자기 자신을 생성하는 **seq2seq**를 학습시켜 얻은
Encoder LSTM의 weights를 **initialization**으로 사용 후 Fine-Tuning

03. Models

✓ Model 1: LSTM

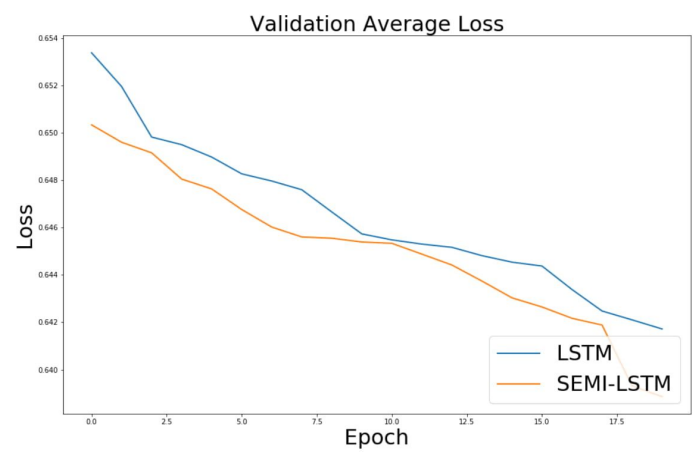
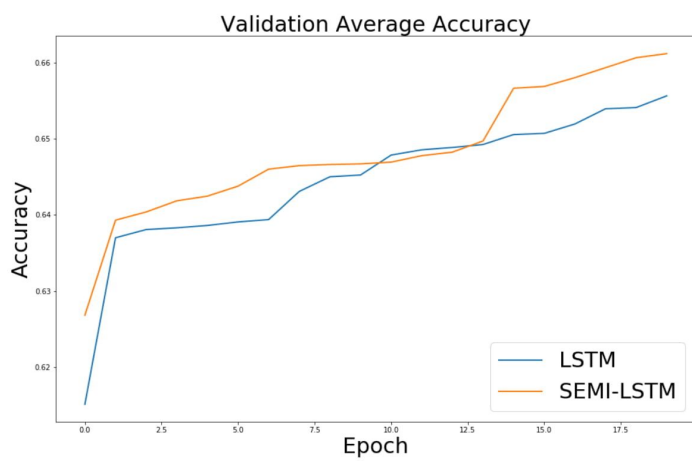
- Initialize weight with pre-trained weight + Add Classifier



03. Models

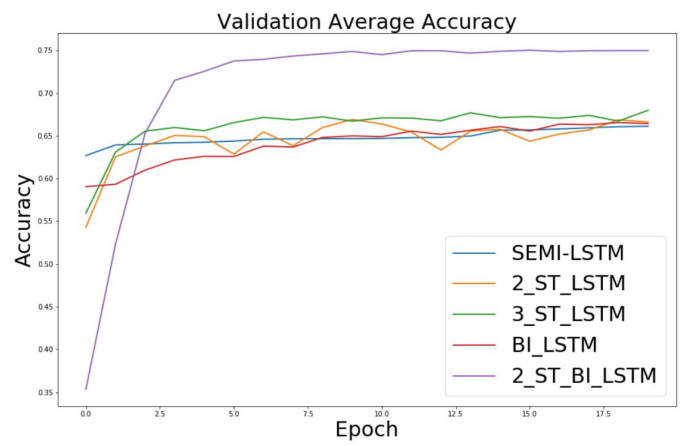
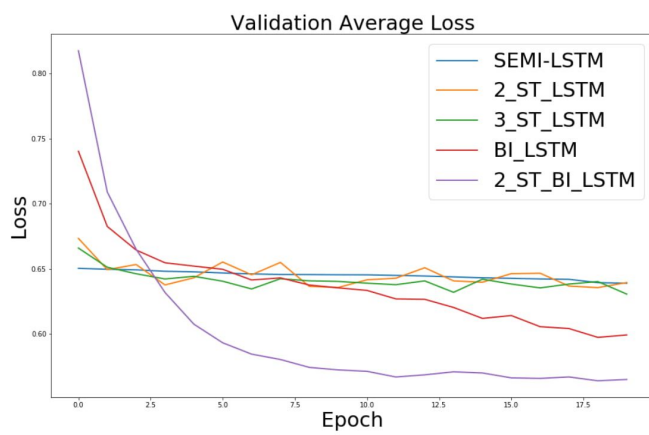
✓ Model 1: LSTM

- Initialize weight with pre-trained weight + Add Classifier



03. Models

✓ Model 2: Deep LSTM Models



03. Models

✓ Model Result

	sec/epoch		test accuracy
	pretrain	train	
LSTM	-	26.5	0.2626
SEMI-LSTM	28.7	26.5	0.2626
2_ST_LSTM	35	28.6	0.2626
3_ST_LSTM	50.4	31.2	0.6883
BI_LSTM	49	28.6	0.7002
2_ST_BI_LSTM	82.6	46.8	0.7362

03. Models

✓ Model 3: BERT

What is BERT?

BERT (**B**idirectional **E**ncoder **R**epresentations from **T**ransformer)



Input
Representation

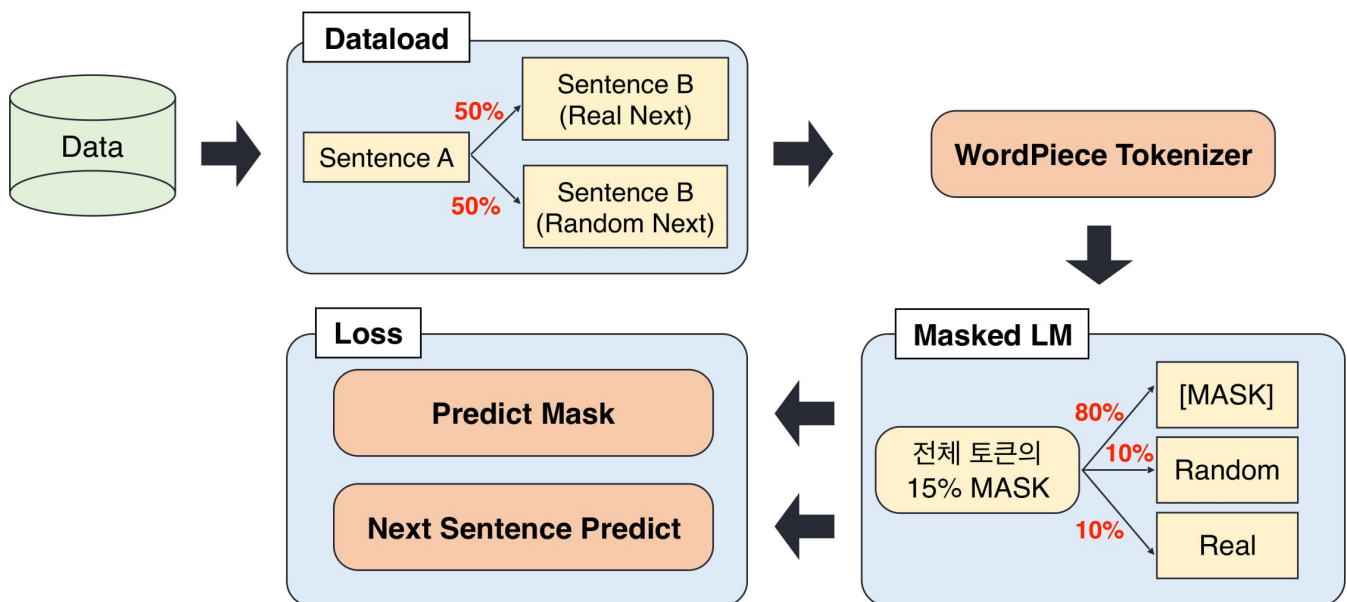
Pretrain

Encoder

03. Models

✓ Model 3: BERT

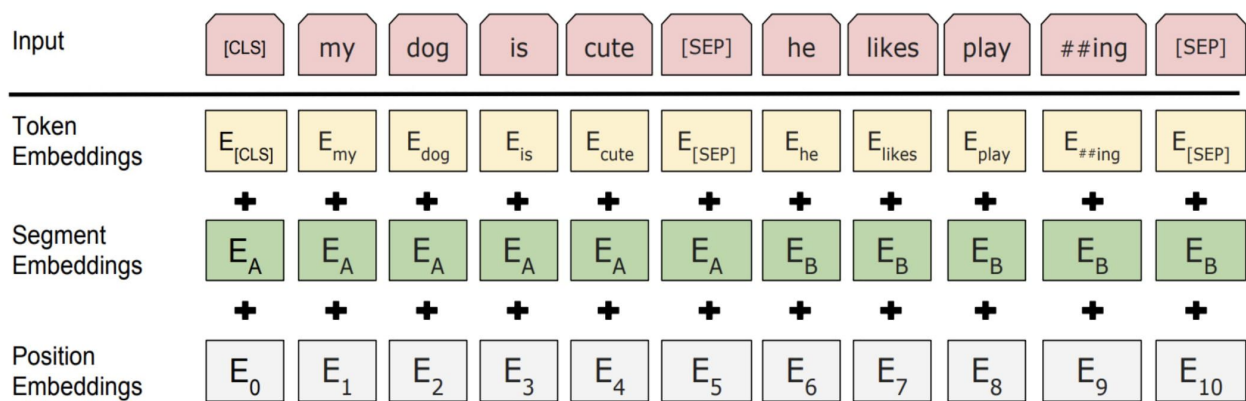
- Pretrain Architecture



03. Models

✓ Model 3: BERT

- Input Representation



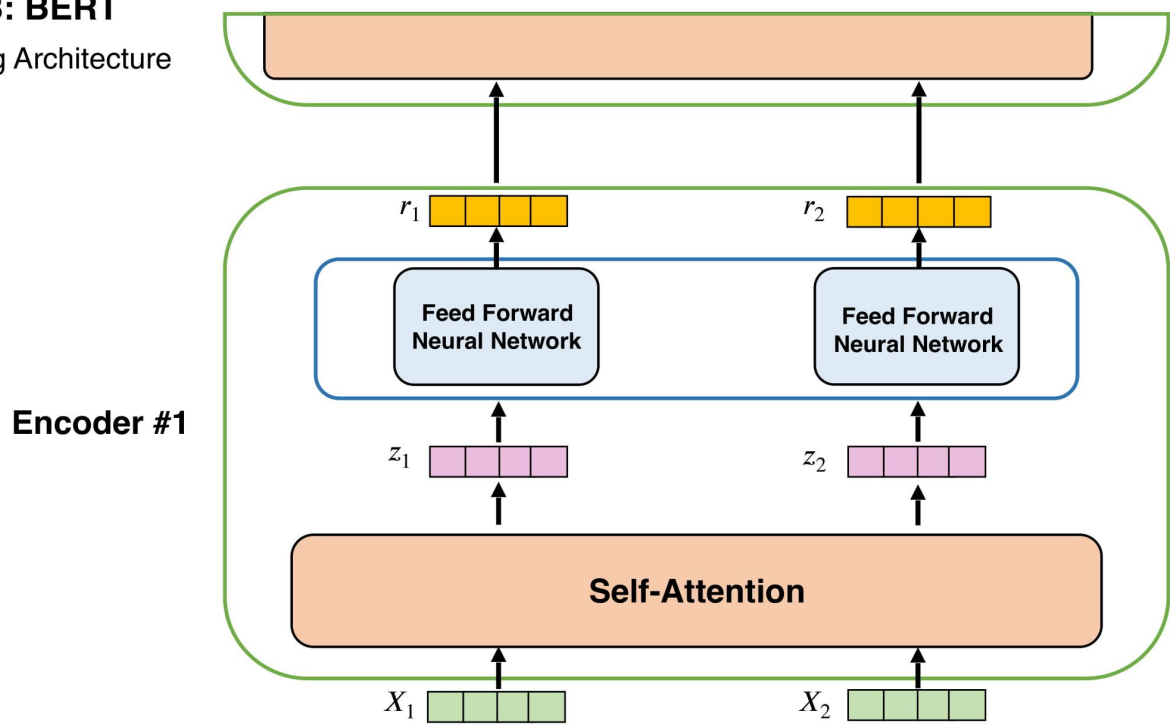
Token Embeddings WordPiece embeddings with a 30,000 token vocabulary.

Segment Embeddings 토큰이 속하는 문장 (A or B)

Position Embeddings 문장 내에서 토큰의 위치

03. Models

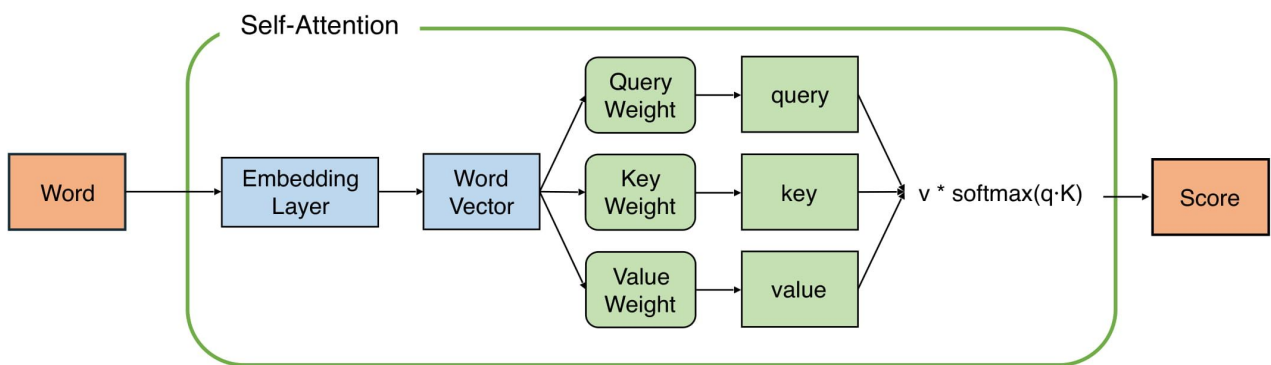
- ✓ **Model 3: BERT**
 - Encoding Architecture



03. Models

✓ Model 3: BERT

- Encoding Self-Attention



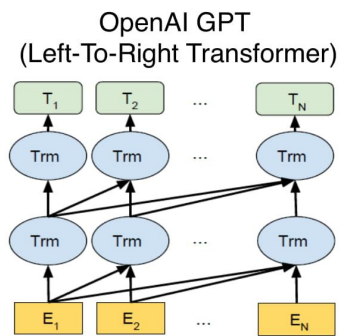
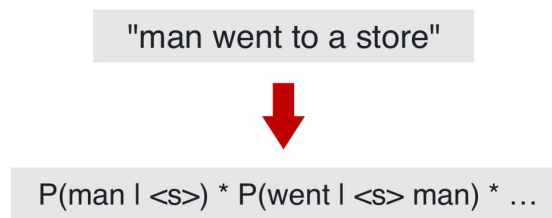
Query에 맞는 최적의 (key, value)값을 찾기

Key: 특정 위치에 단어를 인코딩 할 때 문장의 다른 부분에 얼마나 많이 집중(focus)할지 결정하는 weight

03. Models

✓ Model 3: BERT

- Bidirectional

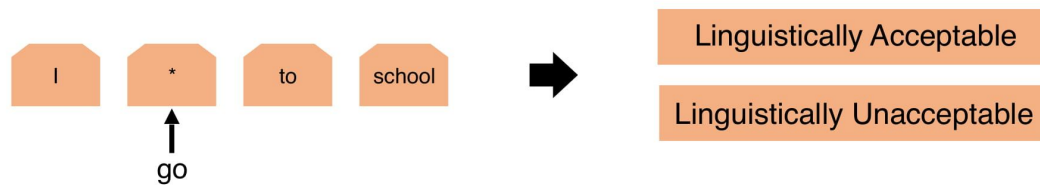


왼쪽 부분의 context만 보기 때문에 한 쪽으로만 bias가 쉽게 걸리게 됨

03. Models

✓ Model 3: BERT

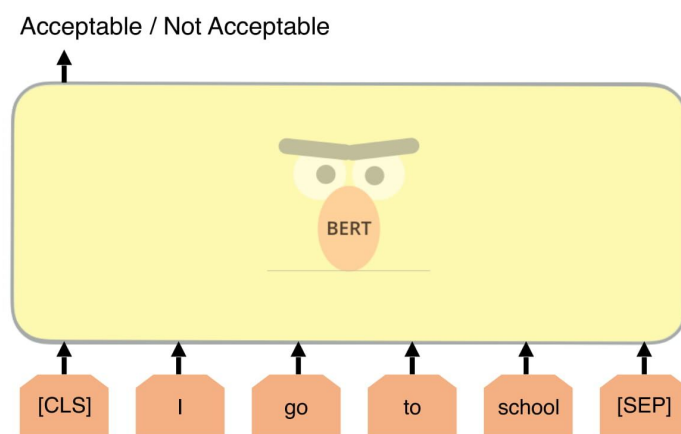
- Data



03. Models

✓ Model 3: BERT

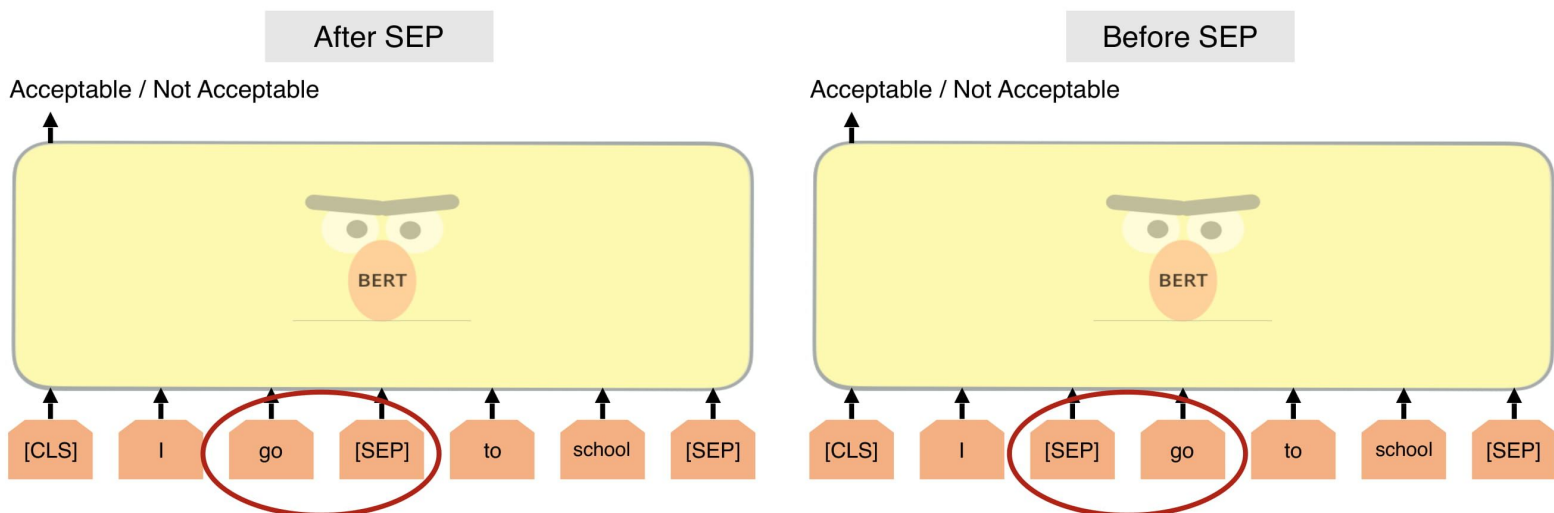
- Single Sentence



03. Models

✓ Model 4: BERT

- Two Sentence



03. Models

✓ Result

	sec/Epoch		Result
	Pretrain	Train	Accuracy
2_ST_BI_LSTM	82.6	46.8	0.7362
BERT_SINGLE	-	86	0.8655
BERT_TWO (Before SEP)	-	85.6	0.8588
BERT_TWO (After SEP)	-	85.2	0.8542

04. Conclusion

✓ 2019학년도 대학수학능력시험 영어 영역

30. 다음 글의 밑줄 친 부분 중, 문맥상 낱말의 쓰임이 적절하지 않은 것은?

Europe's first Homo sapiens lived primarily on large game, particularly reindeer. Even under ideal circumstances, hunting these fast animals with spear or bow and arrow is an ① uncertain task. The reindeer, however, had a ② weakness that mankind would mercilessly exploit: it swam poorly. While afloat, it is uniquely ③ vulnerable, moving slowly with its antlers held high as it struggles to keep its nose above water. At some point, a Stone Age genius realized the enormous hunting ④ advantage he would gain by being able to glide over the water's surface, and built the first boat. Once the ⑤ laboriously overtaken and killed prey had been hauled aboard, getting its body back to the tribal camp would have been far easier by boat than on land. It would not have taken long for mankind to apply this advantage to other goods.

	Pred	Real
1. Even under ideal circumstances, hunting these ...	1	1
2. The reindeer, however, had a ② weakness ...	1	1
3. While afloat, it is uniquely ③ vulnerable, ...	1	1
4. At some point, a Stone Age genius ...	0	1
5. Once the ⑤ laboriously overtaken and ...	1	0

04. Conclusion

✓ 2019학년도 대학수학능력시험 영어 영역

29. 다음 글의 밑줄 친 부분 중, 어법상 틀린 것은? [3점]

“Monumental” is a word that comes very close to ① expressing the basic characteristic of Egyptian art. Never before and never since has the quality of monumentality been achieved as fully as it ② did in Egypt. The reason for this is not the external size and massiveness of their works, although the Egyptians admittedly achieved some amazing things in this respect. Many modern structures exceed ③ those of Egypt in terms of purely physical size. But massiveness has nothing to do with monumentality. An Egyptian sculpture no bigger than a person’s hand is more monumental than that gigantic pile of stones ④ that constitutes the war memorial in Leipzig, for instance. Monumentality is not a matter of external weight, but of “inner weight.” This inner weight is the quality which Egyptian art possesses to such a degree that everything in it seems to be made of primeval stone, like a mountain range, even if it is only a few inches across or ⑤ carved in wood.

	Pred	Real
1. “Monumental” is a word that comes very close...	1	1
2. Never before and never since has the quality of...	0	0
3. Many modern structures exceed ③ those of ...	1	1
4. An Egyptian sculpture no bigger than ...	1	1
5. This inner weight is the quality which Egyptian...	1	1

04. Conclusion

✓ 한계점

TOEIC & CoLA dataset : 문법 문제

- ➡ 수능에서 **문법** 문제 풀이가 가능 하지만,
앞뒤 **맥락**을 통해 풀어야하는 등 학습되지 않은 종류의 문제 풀이 불가능
- ➡ 추후 관련 데이터가 확보된다면 보완 가능

29. 다음 글의 밑줄 친 부분 중,
어법상 틀린 것은?

30. 다음 글의 밑줄 친 부분 중, **문맥상**
낱말의 쓰임이 적절하지 않은 것은?

THANK YOU!