

Machine Reading Comprehension

SSL SEMINAR

20190431 박규현

목차

1 MRC

2 Dataset

3 Model

4문제점

01 _____

MRC : Machine Reading Comprehension

Question answering의 한 종류

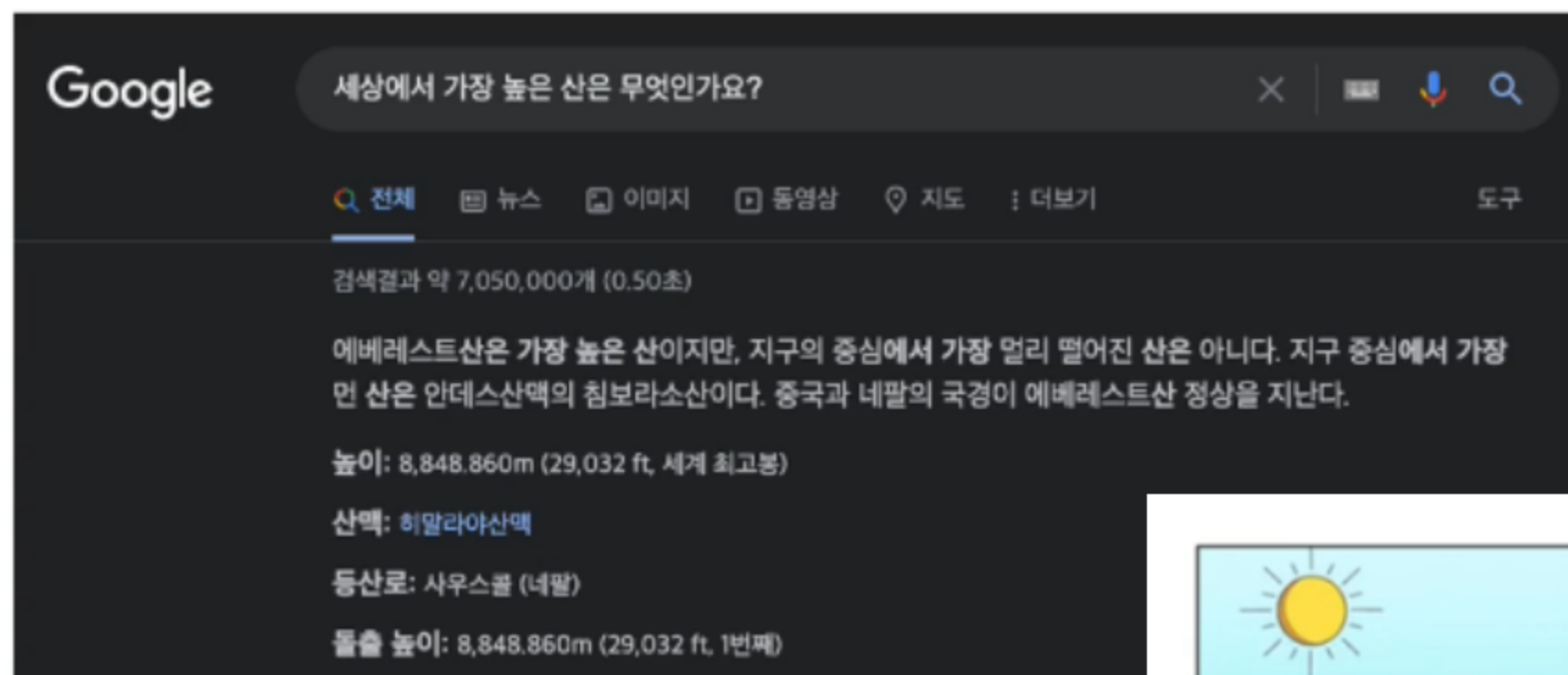


자연어로 표현된 질문에 자동으로 정답을 도출해주는 시스템

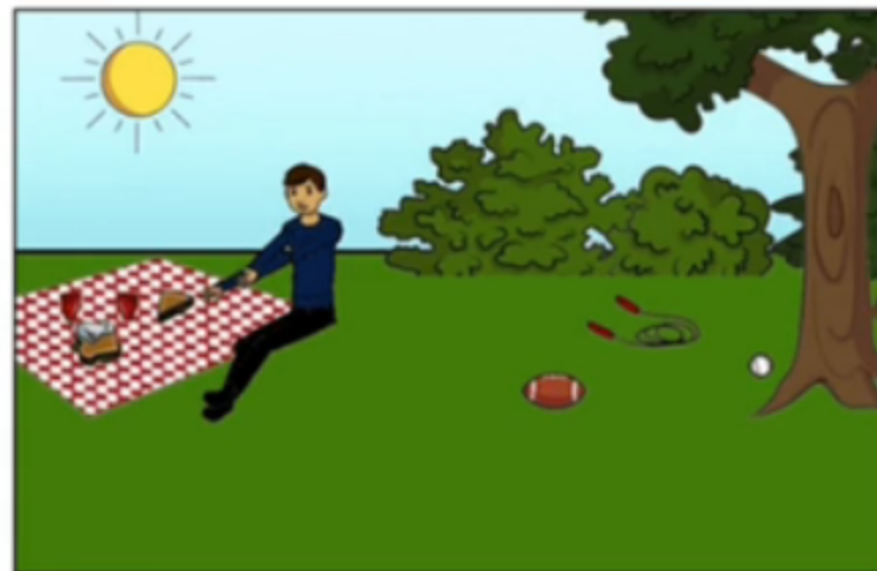
질문의 종류, 답을 찾을 수 있는 source에 따라서 다양한 종류의 Question answering 방법이 가능

01

MRC : Machine Reading Comprehension



OpenDomain Question Answering



Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?

MRC : Machine Reading Comprehension

Reading Comprehension

: 주어진 문단을 이해하고, 질문에 대한
답을 찾는 것

같은 의미인 단어가 존재하면 여러개의
답도 가능함!

Type	Korean
Passage	트라야누스 황제 시대(98년-117년)에 로마 제국은 북으로는 스코틀랜드에서 남으로는 아프리카 수단까지, 서로는 포르투갈의 대서양 연안에서 동으로는 카프카스 지방까지 최대 판도를 이룩했다. 오늘날 면적으로 환산하면 현재 미국 면적의 2/3에 달하고 인구도 현 미국의 절반에 약간 안되는 정도로 추산된다. 서기 5세기 경 서로마 제국은 멸망 후 게르만족의 여러 독립 국가로 갈라져 프랑크 왕국, 신성 로마 제국 등 로마의 후계자를 자처하는 여타 서유럽의 정치 세력들이 나타난다.
Question	로마 제국의 면적이 현재 미국 면적의 2/3에 다다르던 시기는 언제인가?
Answer	트라야누스 황제 시대(98년-117년) / 트라야누스 황제 시대 / 98년-117년

Stanford Question Answering Dataset (SQuAD)

- ◎ 100k 개의 (문단, 질문, 답) 의 쌍이 존재
- ◎ 정답은 문단의 일부분으로 만들어져 있음
- ◎ 정답이 없을 경우, No answer 의 선택지가 있음
- ◎ 가장 유명한 reading comprehension dataset

```
'answers': [{'start': 125, 'end': 128}]}
```

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

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

Dataset Stanford Question Answering Dataset (SQuAD)

◎ Evaluation : exact match (0 or 1) and F1 (partial credit)

Q: What did Tesla do in December 1878?

A: {left Graz, left Graz, left Graz and severed all relations with his family}

Prediction: {left Graz and served}

Exact match: $\max\{0, 0, 0\} = 0$

F1: $\max\{0.67, 0.67, 0.61\} = 0.67$

문단과 질문에 따라서 다양한 정답이 가능할 수
있기 때문에 3개의 gold answer를 설정함

한국어의 경우, F1 대신 ROUGE를 사용함

Dataset Stanford Question Answering Dataset (SQuAD)

◎ ROUGE를 사용하는 이유 ?

한국어의 특성 상 하나의 단어 안에 정답이 존재하는 경우가 있어서 subword-level 의 범위를 정답으로 고려해야함
이를 반영하기 위해서 기존의 한국어 reading comprehension은 character-level F1을 사용하였음
하지만 F1 은 단순히 글자가 같으면 같다고 판단하기 때문에 잘못 판단하는 경우가 있음

Gold answer : "국한된 범위" , Model prediction : "한국의 위인들"

ROUGE score = 15.38, F1 score = 54.55

03

Model

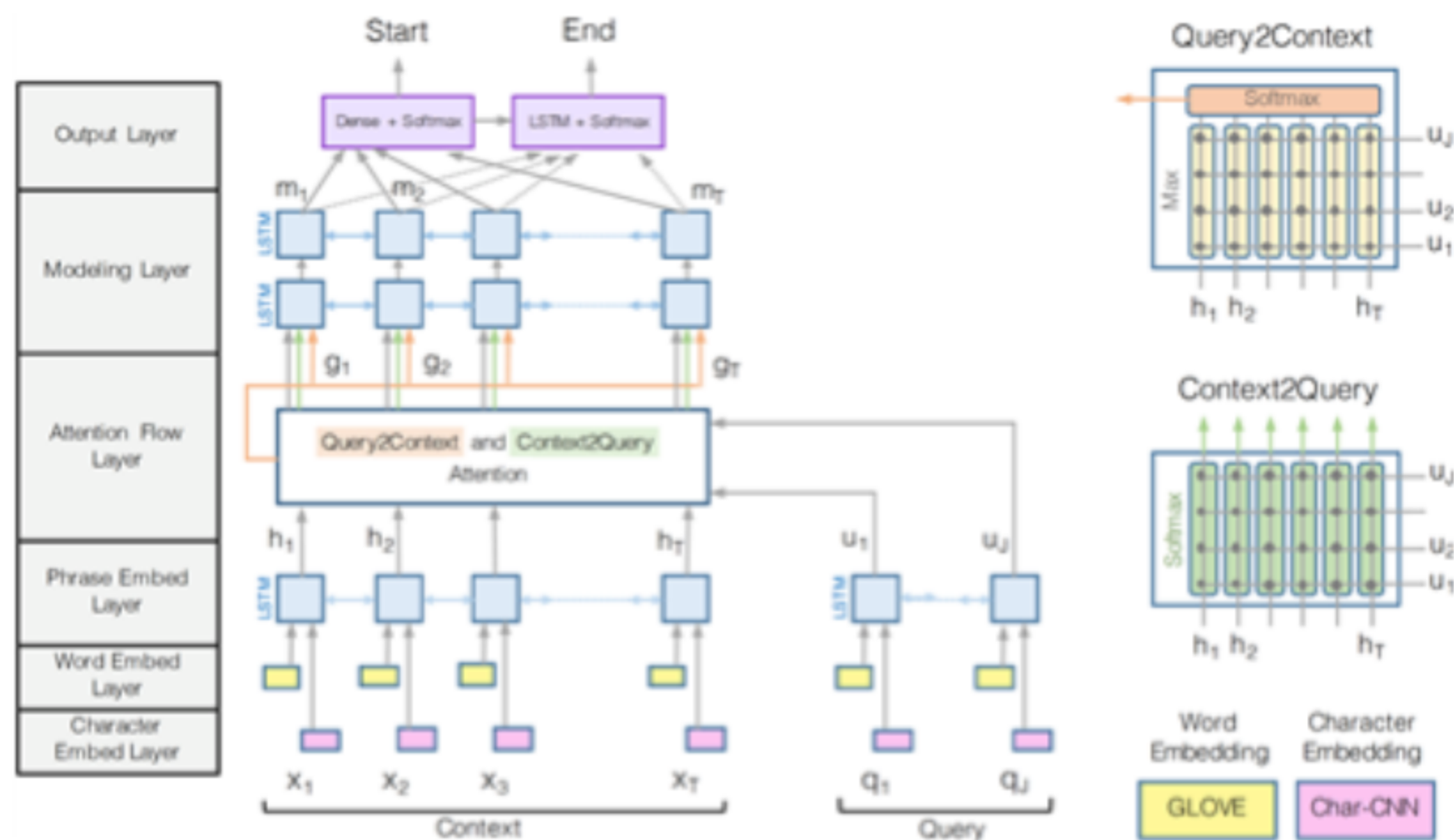
Input : $C = (c_1, c_2, \dots, c_N), Q = (q_1, q_2, \dots, q_M), c_i, q_i \in V$

Output : $1 \leq start \leq end \leq N$

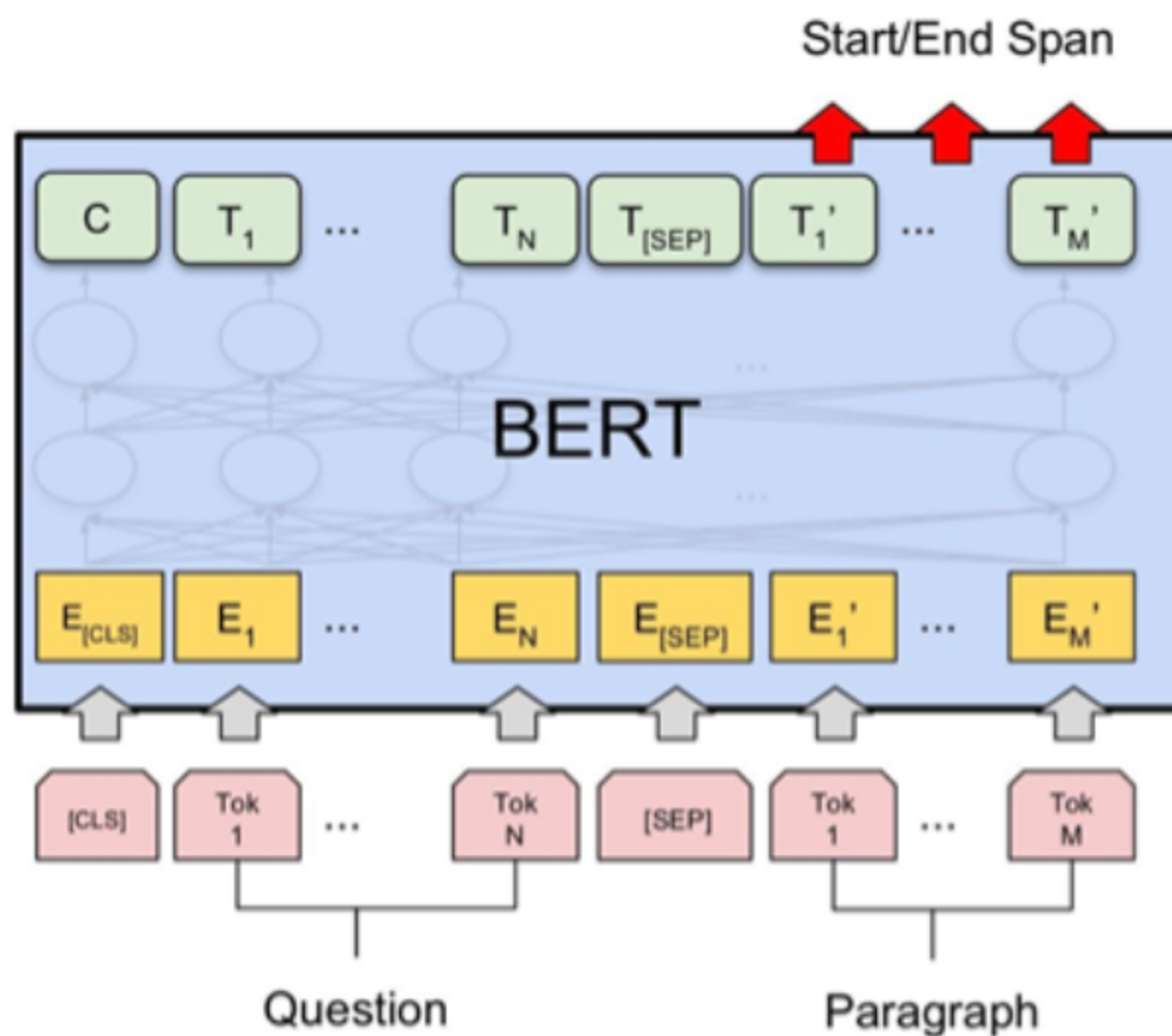
1. LSTM-based models with attention (2016~2018)
 2. Fine-tuning BERT-like models (2019~current)
-

03

Model



LSTM-based



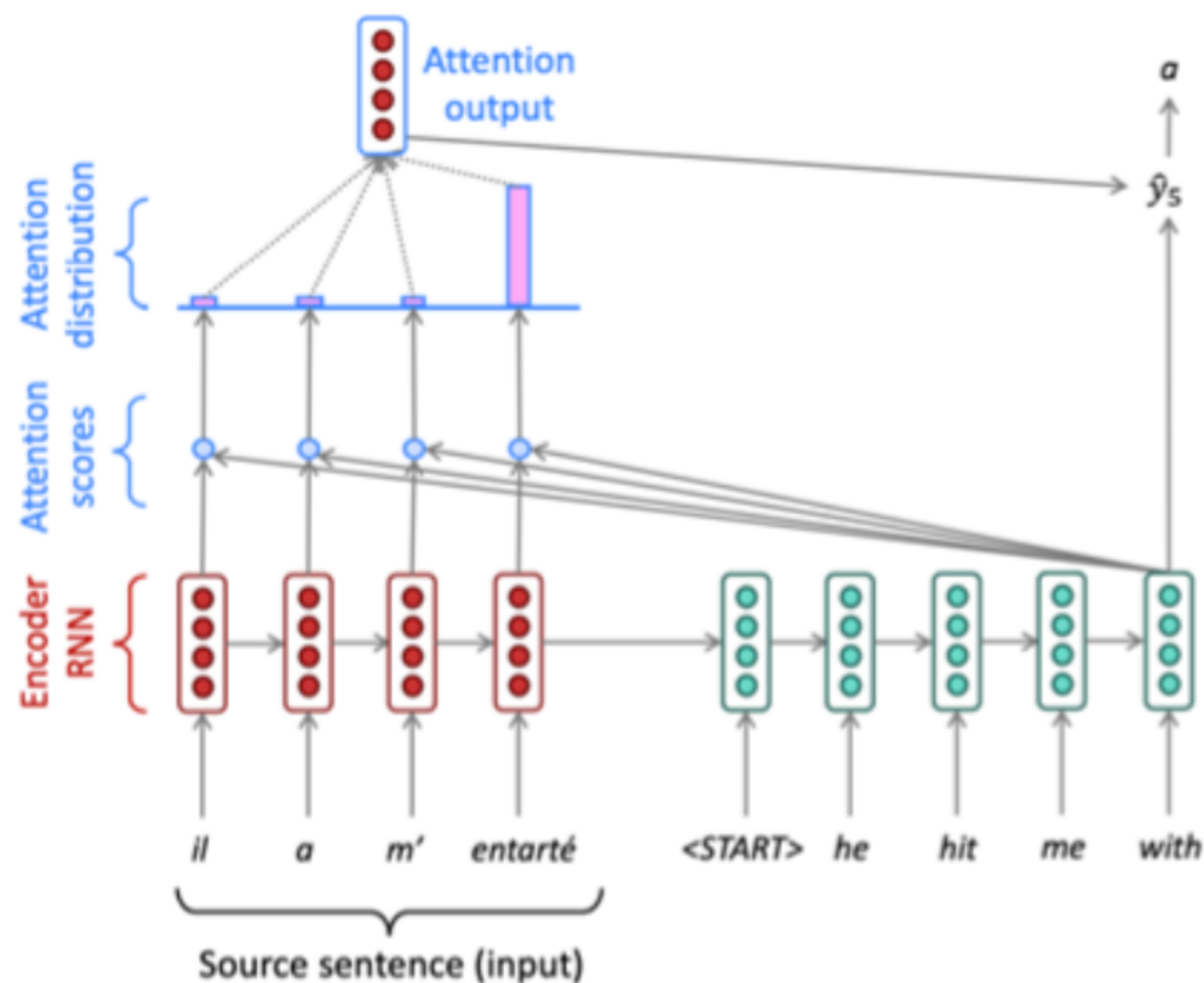
BERT

03

Model : LSTM-based

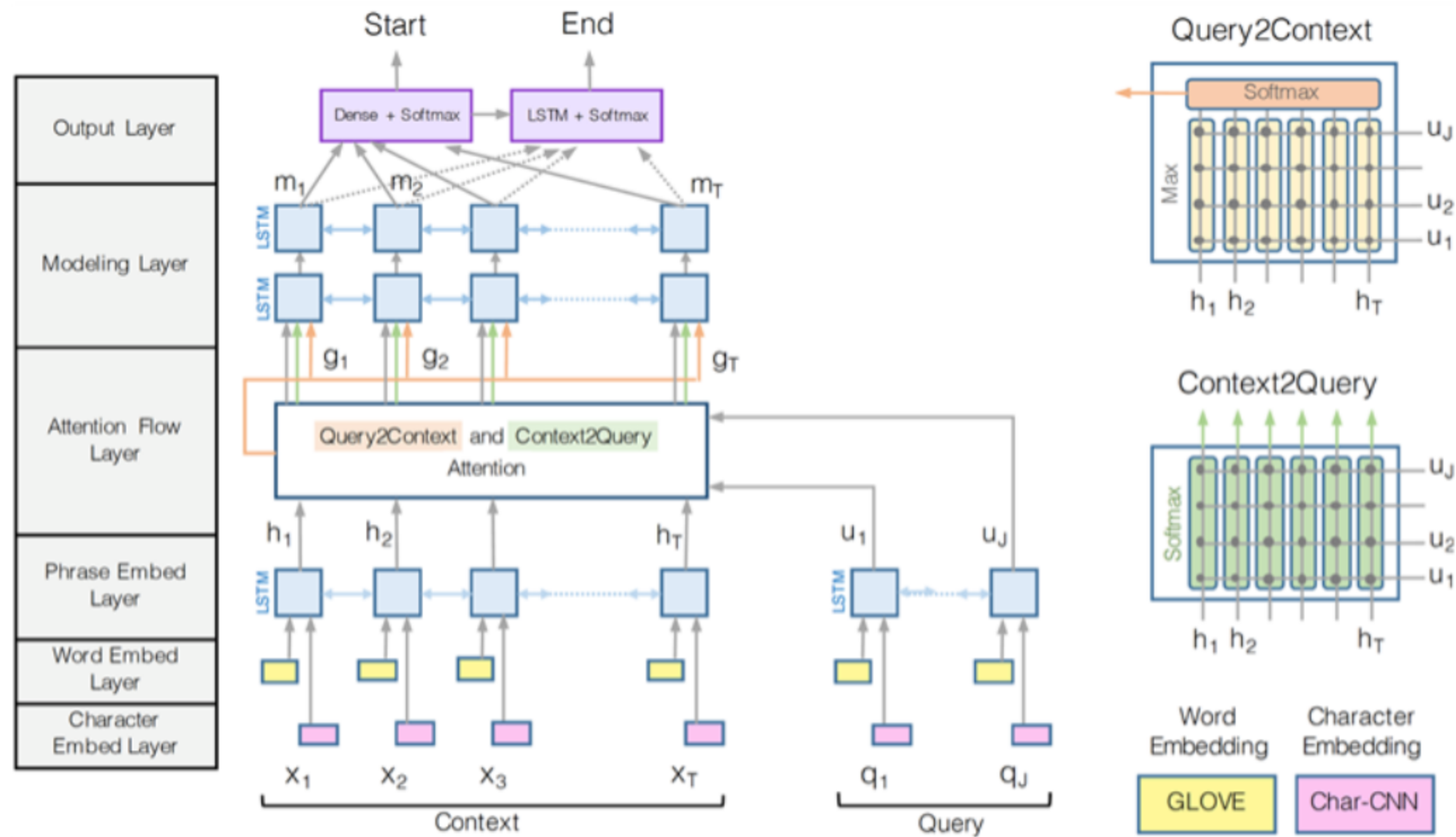
- ◎ Machine translation 에서 사용하는 방법을 reading comprehension에 적용
- ◎ Source, target sentence 대신, passage와 question을 사용
- ◎ Passage 에서 question과 가장 관련이 있는 부분을 찾아야 함
- ◎ 정답의 start, end position을 예측하는 두개의 classifier를 학습함

Seq2seq model with attention



03

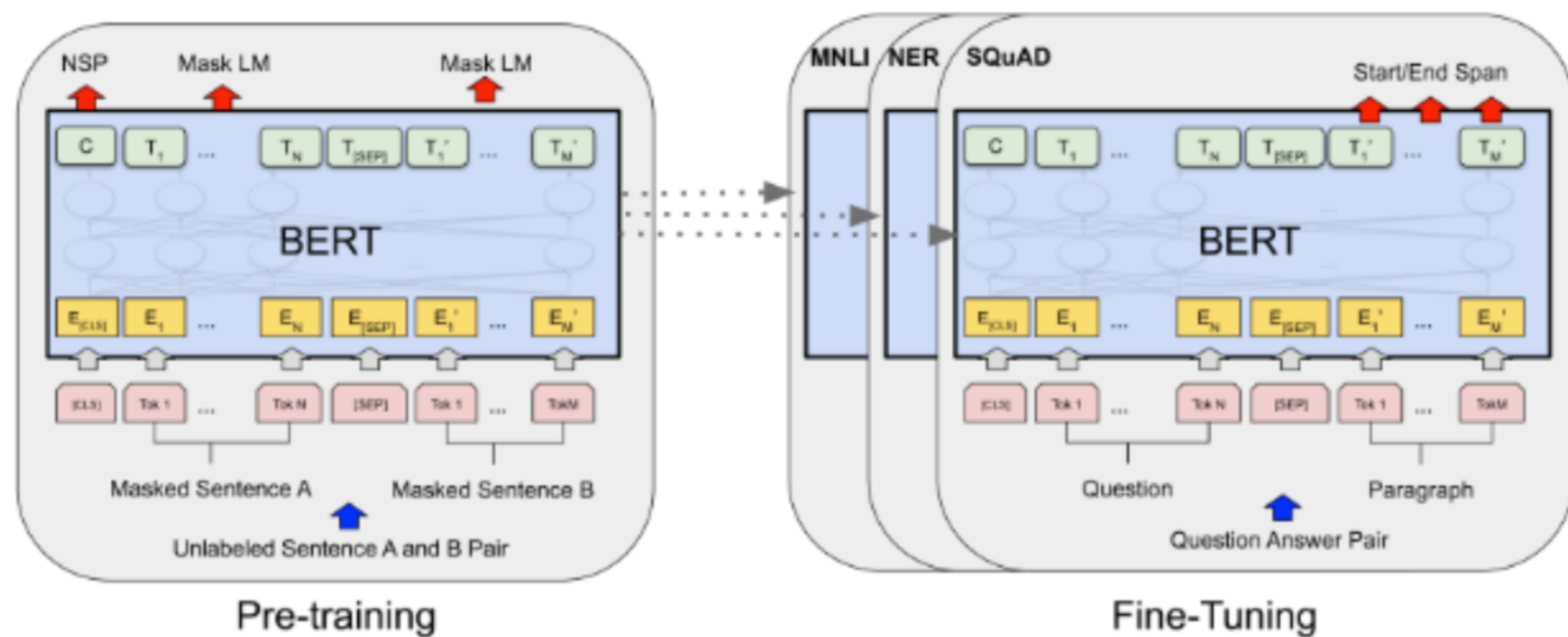
Model : LSTM-based



(Seo et al., 2017) : Bidirectional Attention Flow for Machine Comprehension
 BIDAF: the Bidirectional Attention Flow model

03

Model : BERT



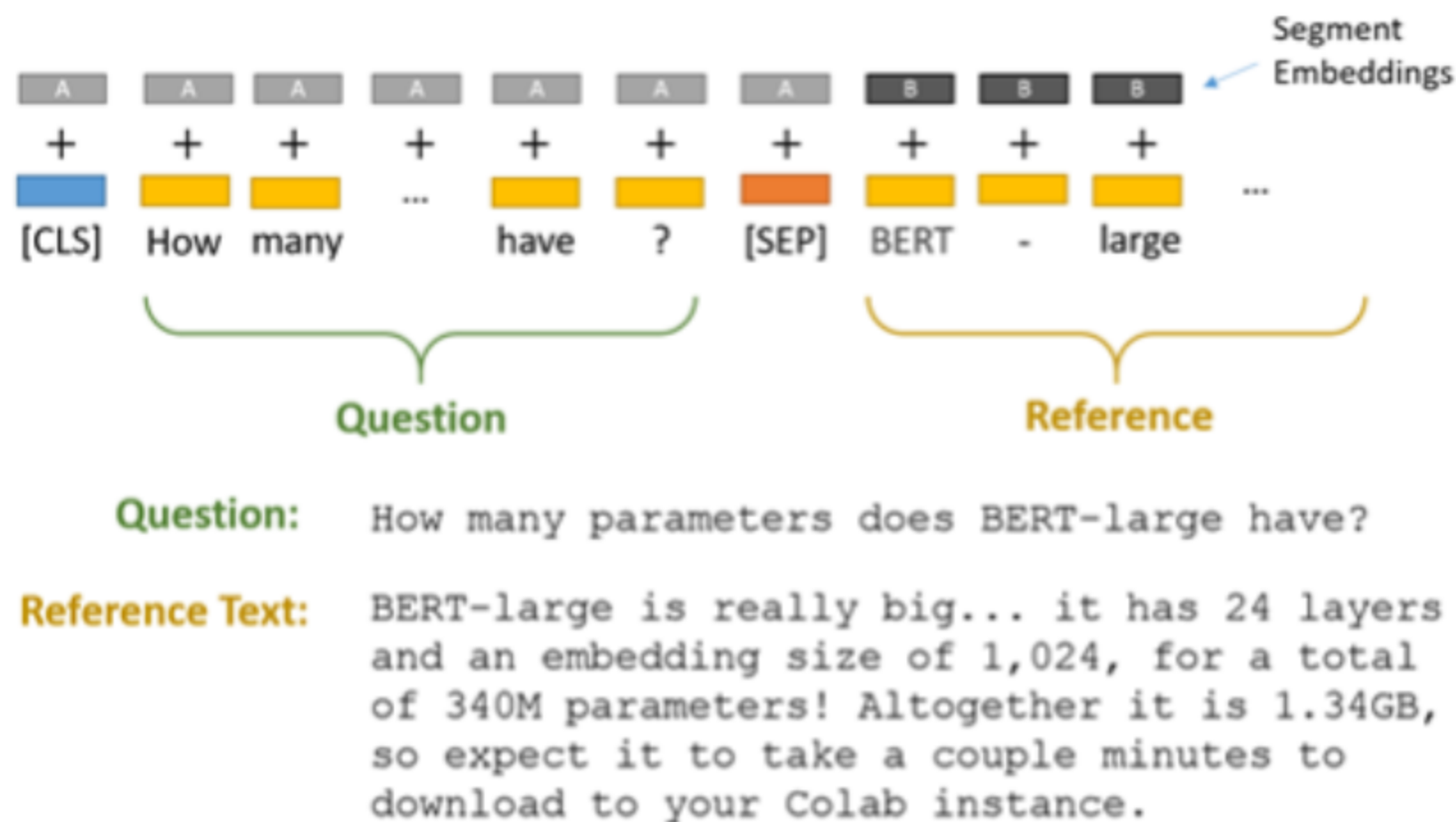
◎ BERT base : 12개의 layer로 구성, 110M개의 파라미터로 구성

◎ BERT large : 24개의 layer로 구성, 330M개의 파라미터로 구성

03

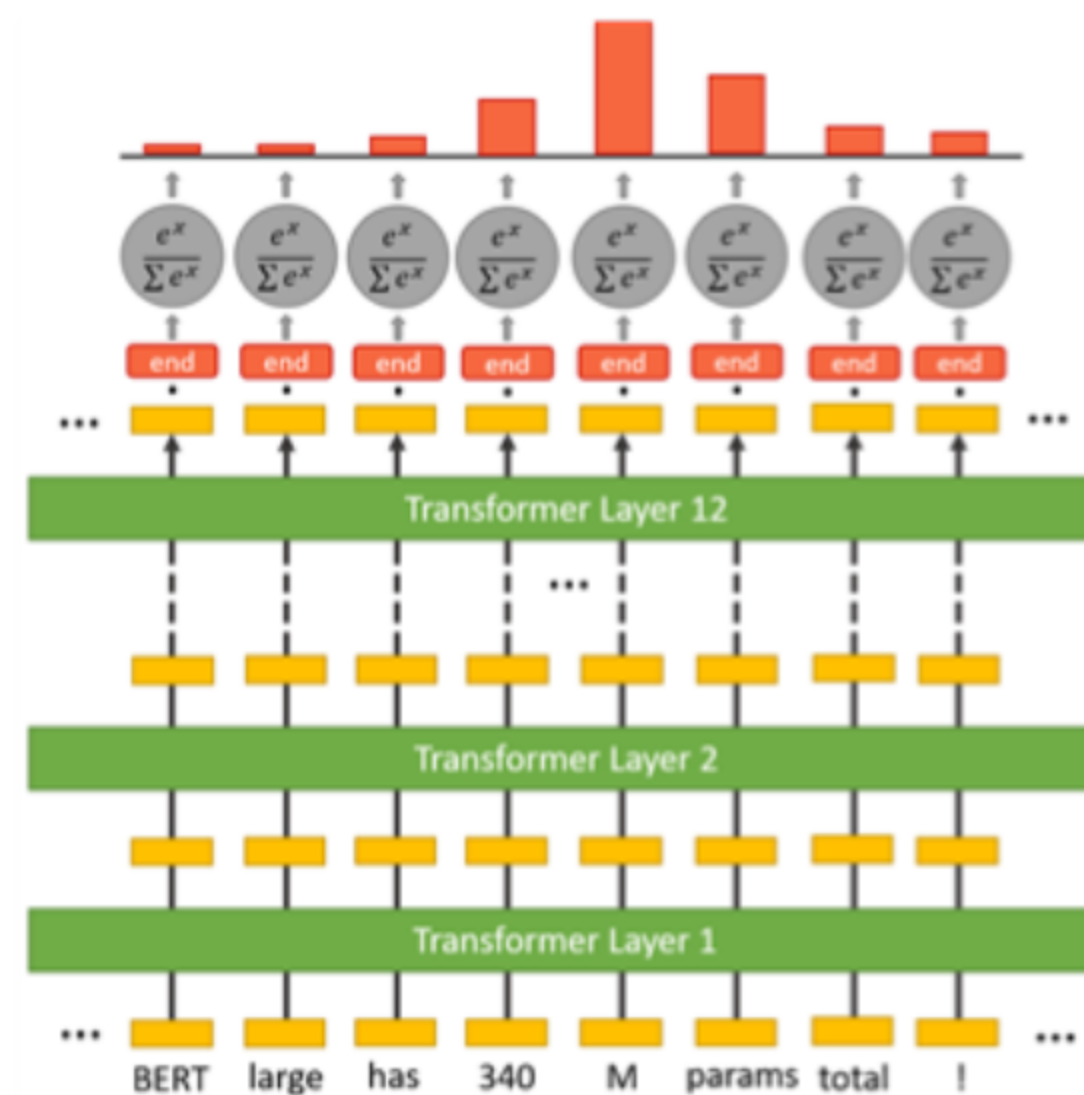
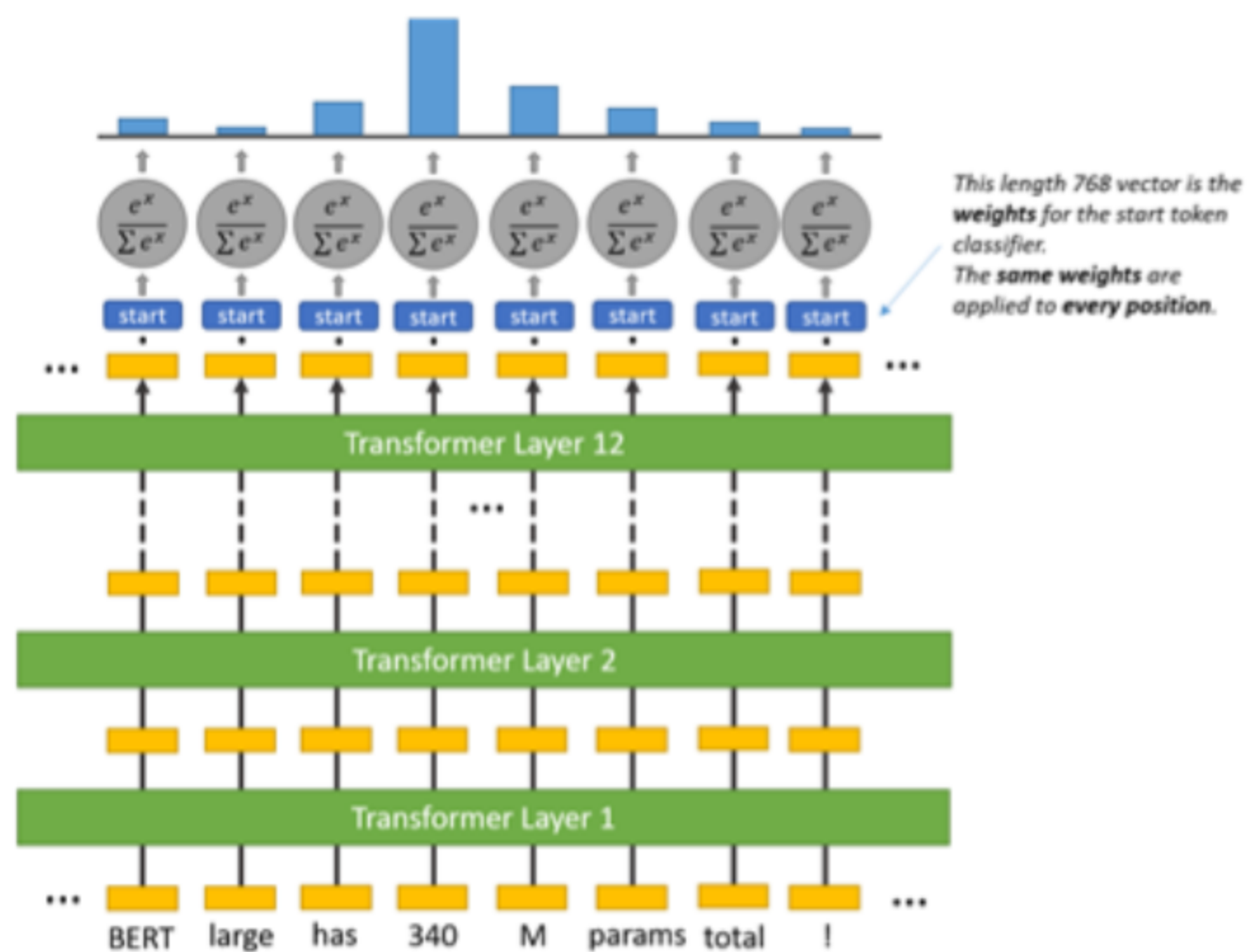
Model : BERT

- ◎ Segment embedding을 사용해서 question과 reference를 구분함
- ◎ [SEP] token 을 사용해서 question과 reference의 경계를 알려줌



03

Model : BERT



03

Model : BERT

	F1	EM
Human performance	91.2*	82.3*
BiDAF	77.3	67.7
BERT-base	88.5	80.8
BERT-large	90.9	84.1
XLNet	94.5	89.0
RoBERTa	94.6	88.9
ALBERT	94.8	89.3

사람의 성능을 뛰어 넘었음

Adversarial examples 들이 추가되었을 때 성능이 확 떨어짐

Article: Super Bowl 50

Paragraph: "Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. *Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.*"

Question: "What is the name of the quarterback who was 38 in Super Bowl XXXIII?"

Original Prediction: John Elway

Prediction under adversary: Jeff Dean

	Match Single	Match Ens.	BiDAF Single	BiDAF Ens.
Original	71.4	75.4	75.5	80.0
ADDSSENT	27.3	29.4	34.3	34.2
ADDONESSENT	39.0	41.8	45.7	46.9
ADDANY	7.6	11.7	4.8	2.7
ADDCOMMON	38.9	51.0	41.7	52.6

(Jia and Liang, 2017): Adversarial Examples for Evaluating Reading Comprehension Systems

다양한 dataset에 대해서 generalize 되지 않음

		Evaluated on				
		SQuAD	TriviaQA	NQ	QuAC	NewsQA
Fine-tuned on	SQuAD	75.6	46.7	48.7	20.2	41.1
	TriviaQA	49.8	58.7	42.1	20.4	10.5
	NQ	53.5	46.3	73.5	21.6	24.7
	QuAC	39.4	33.1	33.8	33.3	13.8
	NewsQA	52.1	38.4	41.7	20.4	60.1

감사합니다!

20190431 박규현