

## Part. 04

# 실전 데이터 분석 프로젝트

## | BERT

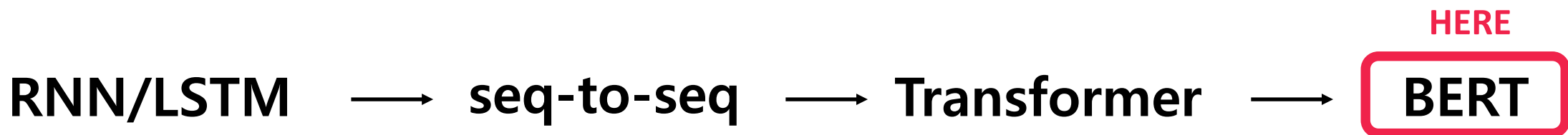
FASTCAMPUS  
ONLINE

머신러닝과 데이터분석A-Z

강사. 김용담

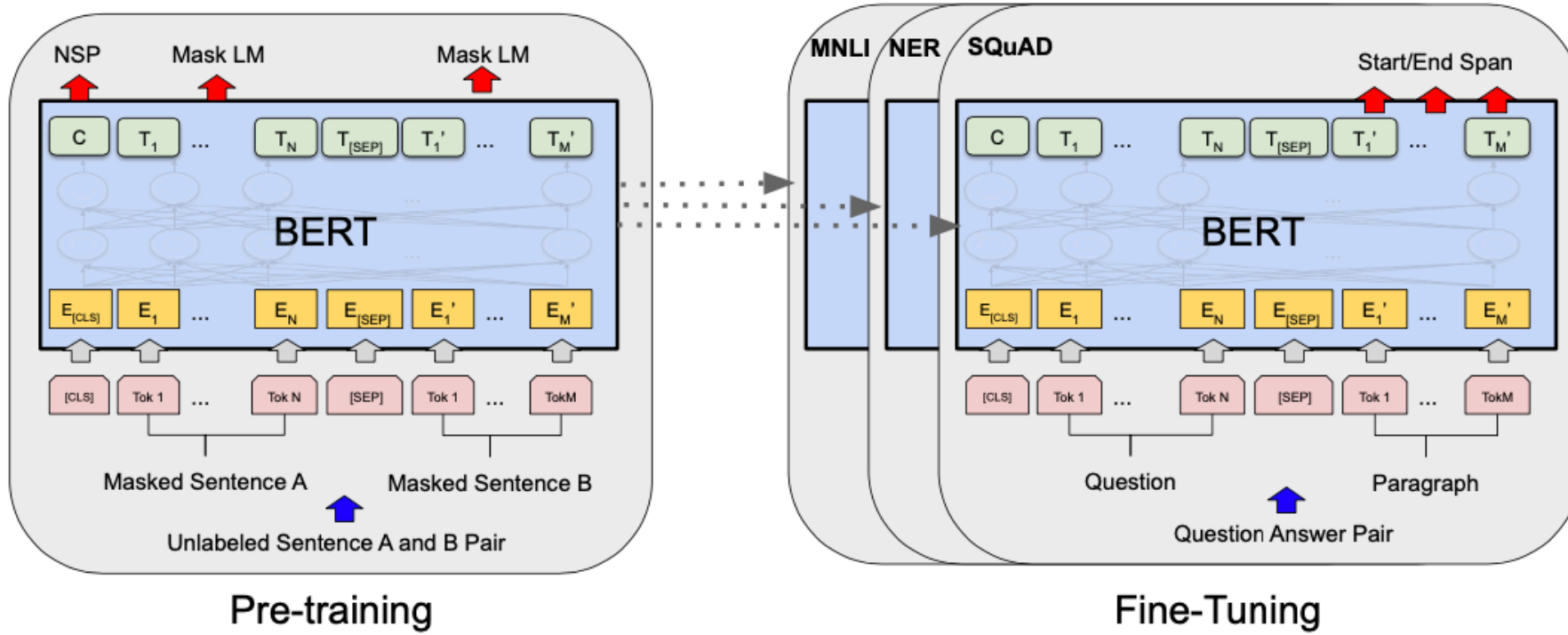
## I NLP Model 역사

## (Bi-directional Encoder for Representation from Transformer)



- |  |  |  |   |
|--|--|--|---|
| <ul style="list-style-type: none"> <li>• 텍스트를 읽는 방법이 이전의 데이터의 정보를 반영하는 것에 착안하여 만들어진 NN 모델</li> <li>• 길이가 길면, 입력이 먼저 들어온 노드일수록 학습에 반영이 되지 않음</li> </ul> | <ul style="list-style-type: none"> <li>• Neural Machine Translation(NMT)의 시대를 열게 된 모델</li> <li>• RNN or LSTM을 두 개를 붙여서 하나를 encoder, 하나를 decoder로 사용하는 모델</li> <li>• 여전히 RNN/LSTM이 길어지면 문제가 생김</li> </ul> | <ul style="list-style-type: none"> <li>• Seq-to-seq의 문제였던 입력이 먼저 들어오는 노드가 정보를 주지 못했던 것을 attention mechanism으로 해결</li> <li>• seq-to-seq에 attention을 추가한 방법을 self-attention이라는 구조로 바꾸어, 계산량을 급격하게 감소시킴</li> <li>• 문장 단위의 embedding quality도 좋아짐</li> </ul> | <ul style="list-style-type: none"> <li>• Pretrained Language Model로 11개의 NLP downstream task의 모든 SOTA 성능을 갈아치우면서 등장</li> <li>• 엄청난 양의 파라미터와 학습량이 필요하지만, Large Model의 경우 성능이 엄청남</li> <li>• 특정 task 뿐만 아니라, 언어에 대한 이해 자체가 목표인 대용량 모델의 등장으로 multi-lingual language model이 많이 생겨남</li> </ul> |
|--|--|--|---|

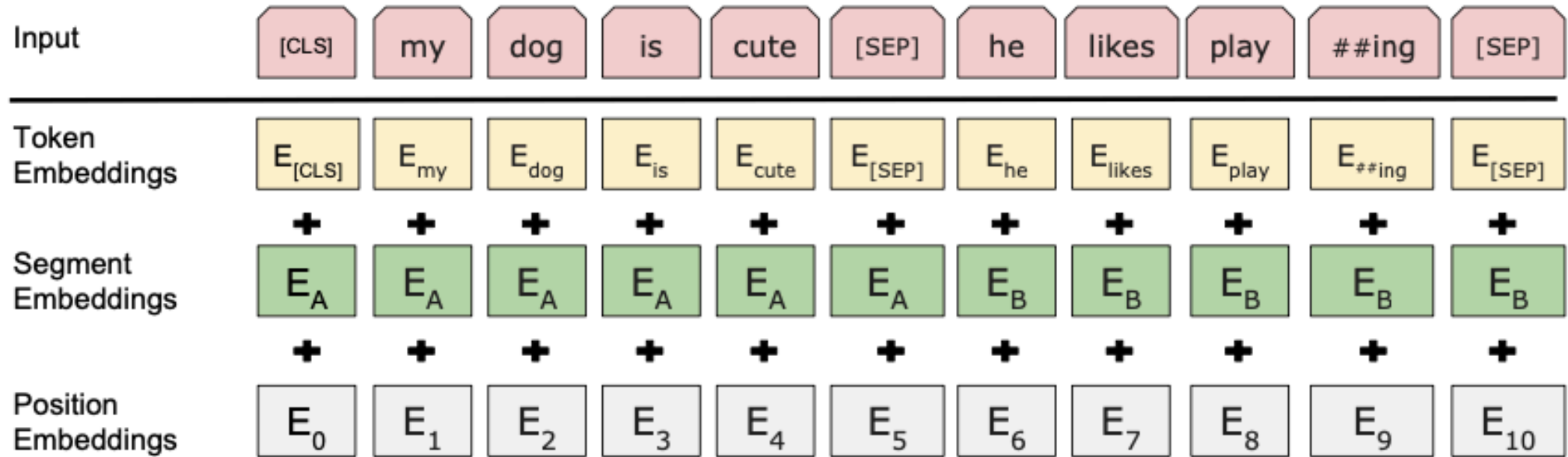
# BERT의 구조



## BERT Structure

Reference : BERT\_Pre-training of Deep Bidirectional Transformers for Language Understanding(NAAACL-HLT 2019)

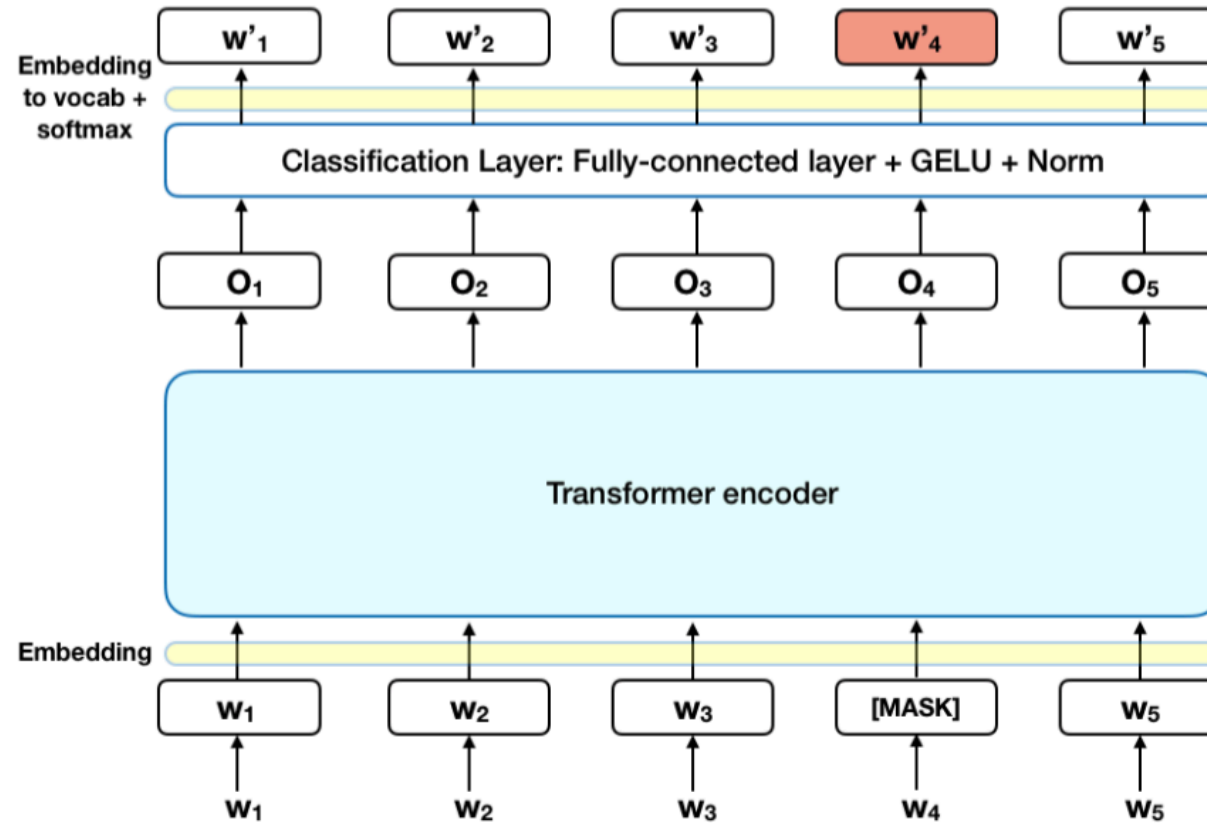
# BERT의 입력



## Input of BERT Model

Reference : BERT\_Pre-training of Deep Bidirectional Transformers for Language Understanding(NAAACL-HLT 2019)

# I Masked LM

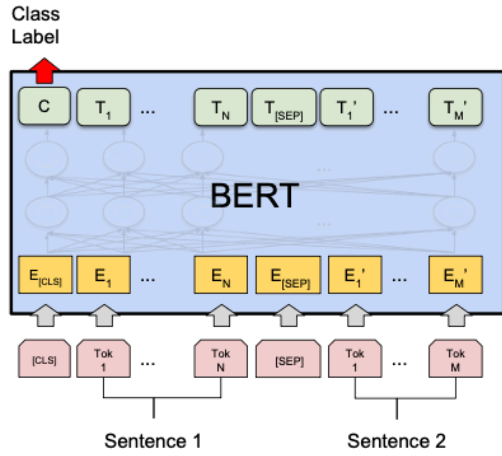


## Masked Language Model

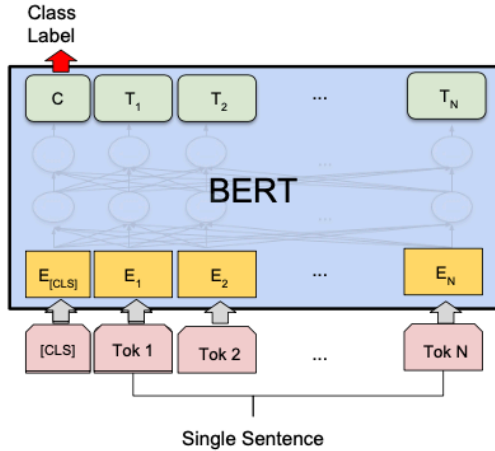
Reference : <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>



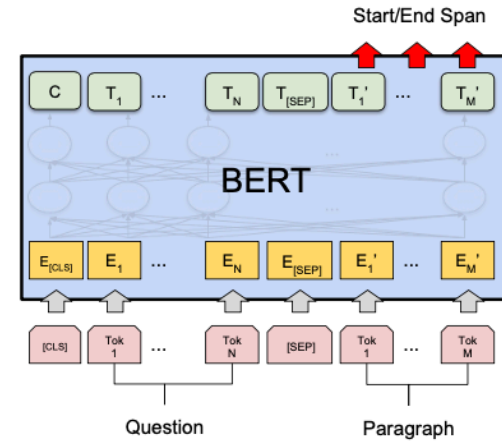
# I BERT의 학습, pre-training, fine-tuning



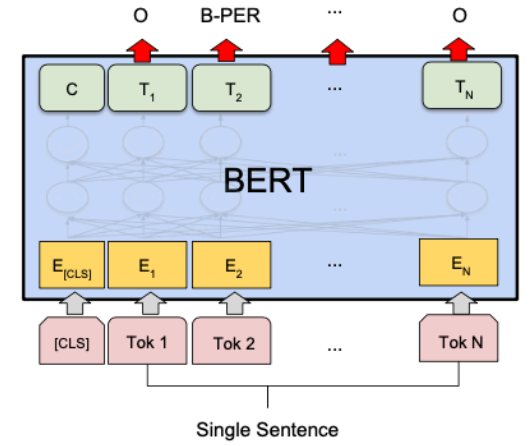
(a) Sentence Pair Classification Tasks:  
MNLI, QQP, QNLI, STS-B, MRPC,  
RTE, SWAG



(b) Single Sentence Classification Tasks:  
SST-2, CoLA



(c) Question Answering Tasks:  
SQuAD v1.1



(d) Single Sentence Tagging Tasks:  
CoNLL-2003 NER

## BERT Fine-tuning Model

Reference : BERT\_Pre-training of Deep Bidirectional Transformers for Language Understanding(NAAACL-HLT 2019)

# I BERT의 성능

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	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
BERT <sub>BASE</sub>	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT <sub>LARGE</sub>	<b>86.7/85.9</b>	<b>72.1</b>	<b>92.7</b>	<b>94.9</b>	<b>60.5</b>	<b>86.5</b>	<b>89.3</b>	<b>70.1</b>	<b>82.1</b>

## Performance of BERT

Reference : BERT\_Pre-training of Deep Bidirectional Transformers for Language Understanding(NAAACL-HLT 2019)

Part. 04

실전 데이터 분석 프로젝트

# I 감사합니다.

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석A-Z

강사. 김용담