

⋮ Hide menu

Question Answering

✓ Video: Week Introduction

41 sec

✓ Video: Week 3 Overview

6 min

✓ Reading: Week 3 Overview

10 min

✓ Video: Transfer Learning in NLP

6 min

✓ Reading: Transfer Learning in NLP

10 min

✓ Video: ELMo, GPT, BERT, T5

7 min

✓ Reading: ELMo, GPT, BERT, T5

10 min

✓ Video: Bidirectional Encoder Representations from Transformers (BERT)

4 min

✓ Reading: Bidirectional Encoder Representations from Transformers (BERT)

10 min

✓ Video: BERT Objective

2 min

✓ Reading: BERT Objective

10 min

✓ Video: Fine tuning BERT

2 min

✓ Reading: Fine tuning BERT

10 min

✓ Video: Transformer: T5

3 min

✓ Reading: Transformer T5

10 min

✓ Video: Multi-Task Training Strategy

5 min

✓ Reading: Multi-Task Training Strategy

10 min

✓ Video: GLUE Benchmark

2 min

✓ Reading: GLUE Benchmark

10 min

✓ Lab: SentencePiece and BPE

2h

Hugging Face

Lecture Notes (Optional)

Practice Quiz

Assignment

Heroes of NLP: Quoc Le

Acknowledgments & Course Resources

# BERT Objective

We will first start by visualizing the input.

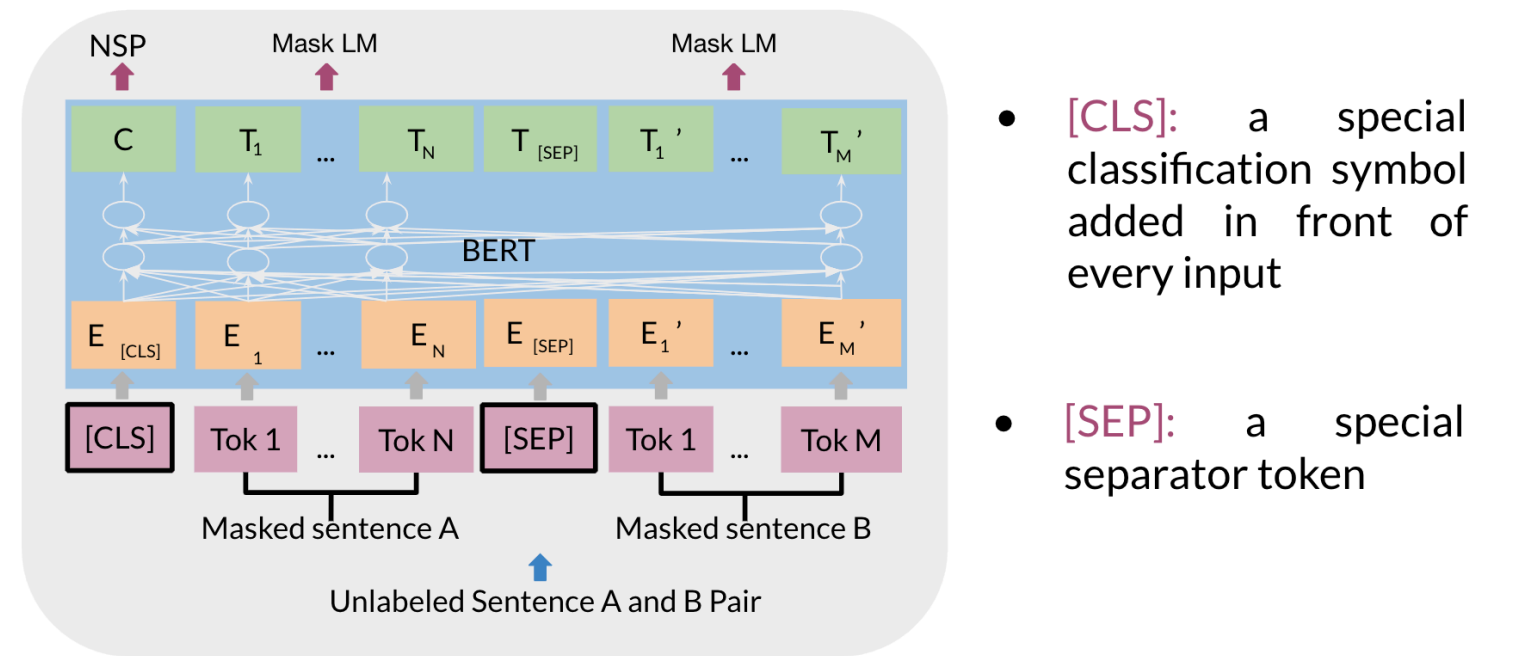
Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	E <sub>[CLS]</sub>	E <sub>my</sub>	E <sub>dog</sub>	E <sub>is</sub>	E <sub>cute</sub>	E <sub>[SEP]</sub>	E <sub>he</sub>	E <sub>likes</sub>	E <sub>play</sub>	E <sub>##ing</sub>	E <sub>[SEP]</sub>
Segment Embeddings	E <sub>A</sub>	E <sub>A</sub>	E <sub>A</sub>	E <sub>A</sub>	E <sub>A</sub>	E <sub>A</sub>	E <sub>B</sub>	E <sub>B</sub>	E <sub>B</sub>	E <sub>B</sub>	E <sub>B</sub>
Position Embeddings	E <sub>0</sub>	E <sub>1</sub>	E <sub>2</sub>	E <sub>3</sub>	E <sub>4</sub>	E <sub>5</sub>	E <sub>6</sub>	E <sub>7</sub>	E <sub>8</sub>	E <sub>9</sub>	E <sub>10</sub>

The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

**The input embeddings:** you have a CLS token to indicate the beginning of the sentence and a sep to indicate the end of the sentence

**The segment embeddings:** allows you to indicate whether it is sentence a or b.

**Positional embeddings:** allows you to indicate the word's position in the sentence.



The C token in the image above could be used for classification purposes. The unlabeled sentence A/B pair will depend on what you are trying to predict, it could range from question answering to sentiment. (in which case the second sentence could be just empty). The BERT objective is defined as follows:

- Objective 1:

Multi-Mask LM

Loss: Cross Entropy Loss
- Objective 2:

Next Sentence Prediction

Loss: Binary Loss

○

○

○

○

○

}

V

+

○

○

○

}

2

You just combine the losses!

If you are interested in digging deeper into this topic, we recommend you to look at [this article](#) ↗.

