

BERT:

B: Bidirectional

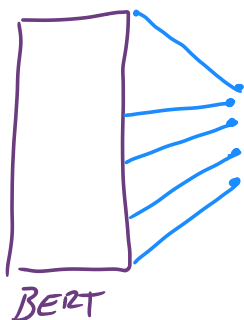
E: Encoder

R: Representations from

T: Transformers

→ Pretrain on unlabeled text (Unsupervised)

→ Joining left and right contexts.



We only need one more layer to use this network for different tasks

Well to: → Named entity recognition
→ Question answering

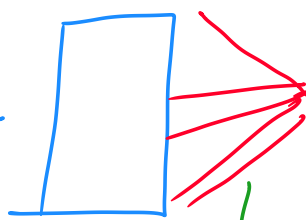
There are two existing strategies for applying pre-trained language representations.

1) Feature based: - Using the pre-trained representations as additional features.

2) Fine-tuning: - Introduces minimal task-specific parameters and is trained on the downstream tasks by simply fine-tuning all pre-trained parameters.

It's very important consider context from both sides.

TRANSFER LEARNING



Retrain = Fine tuning

Feature extraction

BERT: two steps:

1) Pre-training

2) Fine-tuning

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	E_{my}	E_{dog}	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	E_{likes}	E_{play}	$E_{\#ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}