

Datawhale

Bert原理与技巧

对于CV和NLP，预训练模型都很重要：

- ✓词向量具体的形式；
- ✓预训练预料的组成；

对于NLP预训练模型更加重要：

- ✓同一个单词有不同的含义；
- ✓上下文语句不同单词含义不同；

✓学术派：语言模型

✓实践派：从任务进行学习

✓具体有什么任务，任务的标注，常见的模型是什么？

1-of-N Encoding

apple = [1 0 0 0 0]

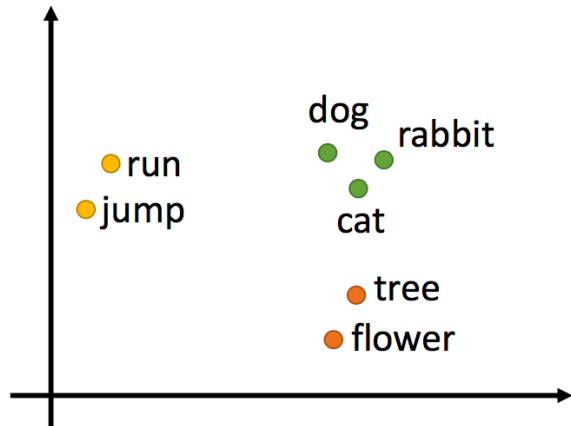
bag = [0 1 0 0 0]

cat = [0 0 1 0 0]

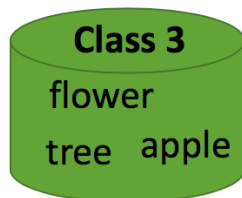
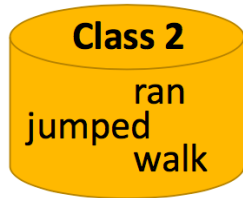
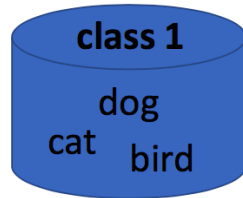
dog = [0 0 0 1 0]

elephant = [0 0 0 0 1]

Word Embedding



Word Class



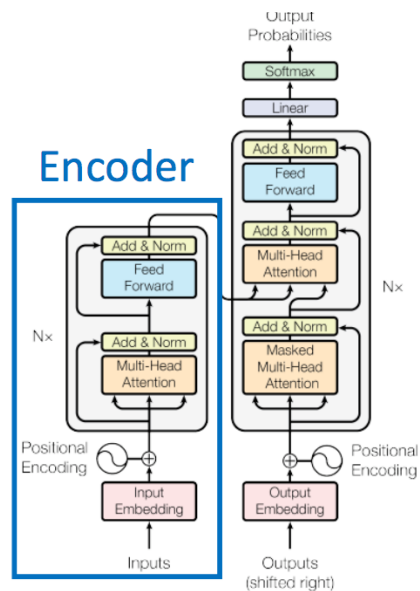
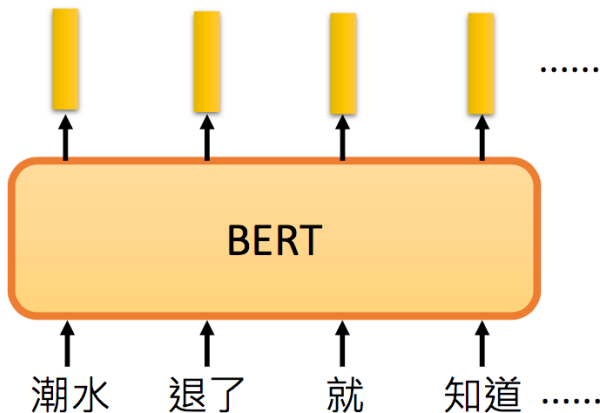
□ BERT (Bidirectional Encoder Representations from Transformers)

<https://arxiv.org/abs/1810.04805>

✓Transformer中的Encoder就是Bert预训练的架构

- BERT = Encoder of Transformer

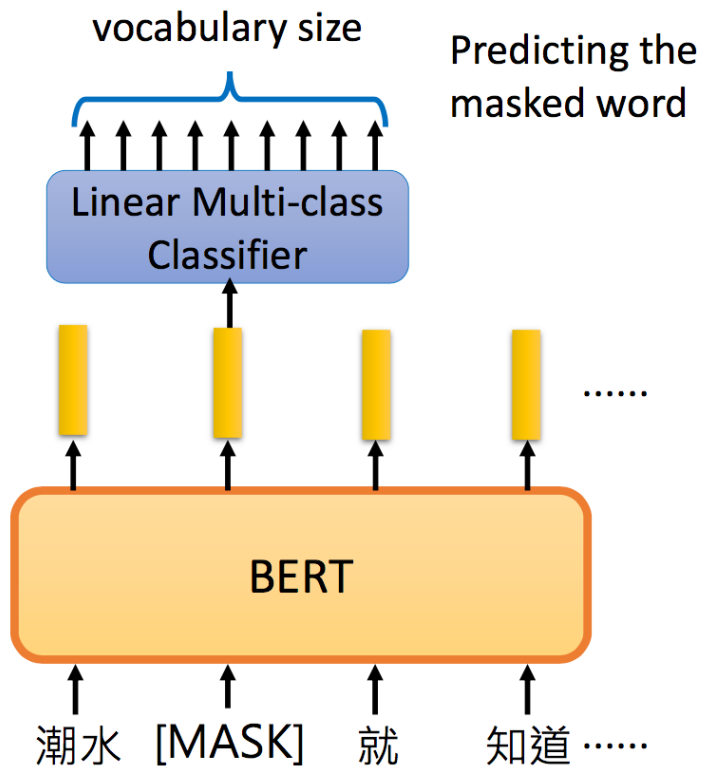
Learned from a large amount of text
without annotation



□ BERT (Bidirectional Encoder Representations from Transformers)

Masked LM: 随机把一些单词变为Mask, 让模型去猜测盖住的地方是什么单词。

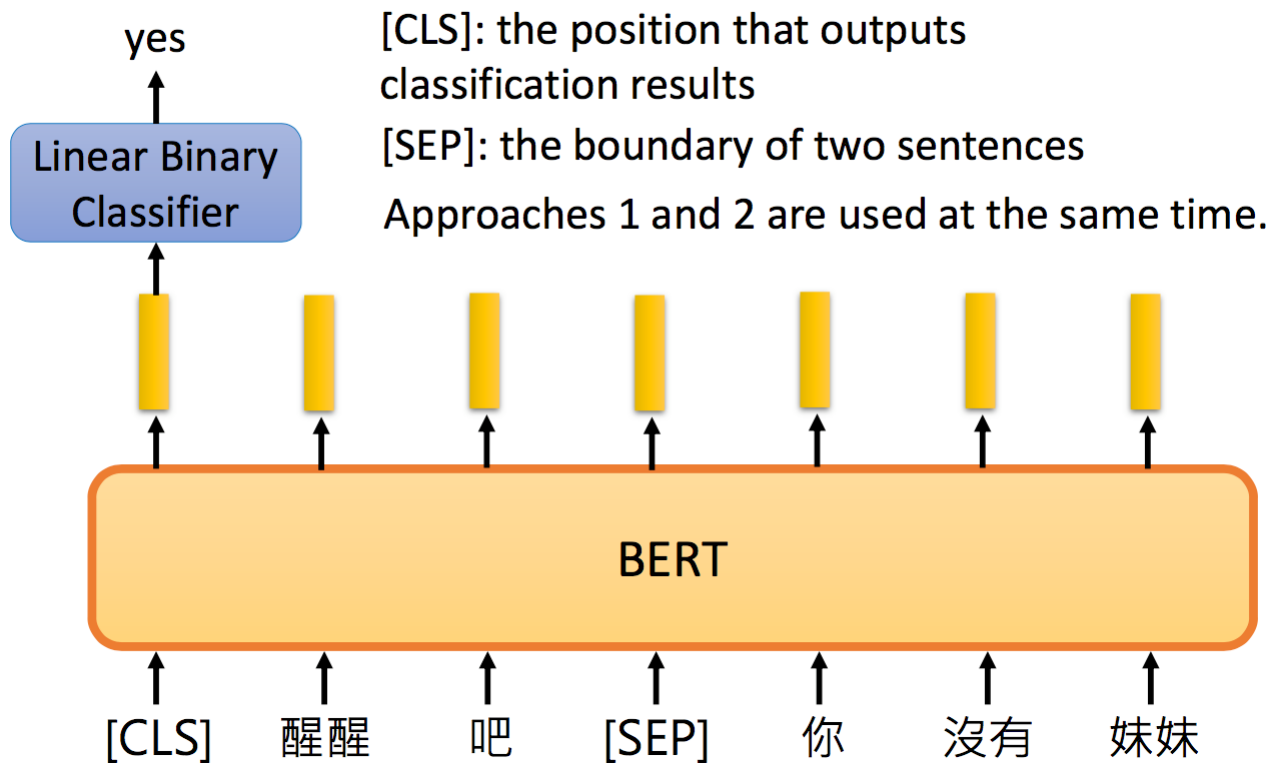
- Approach 1:
Masked LM



□ BERT (Bidirectional Encoder Representations from Transformers)

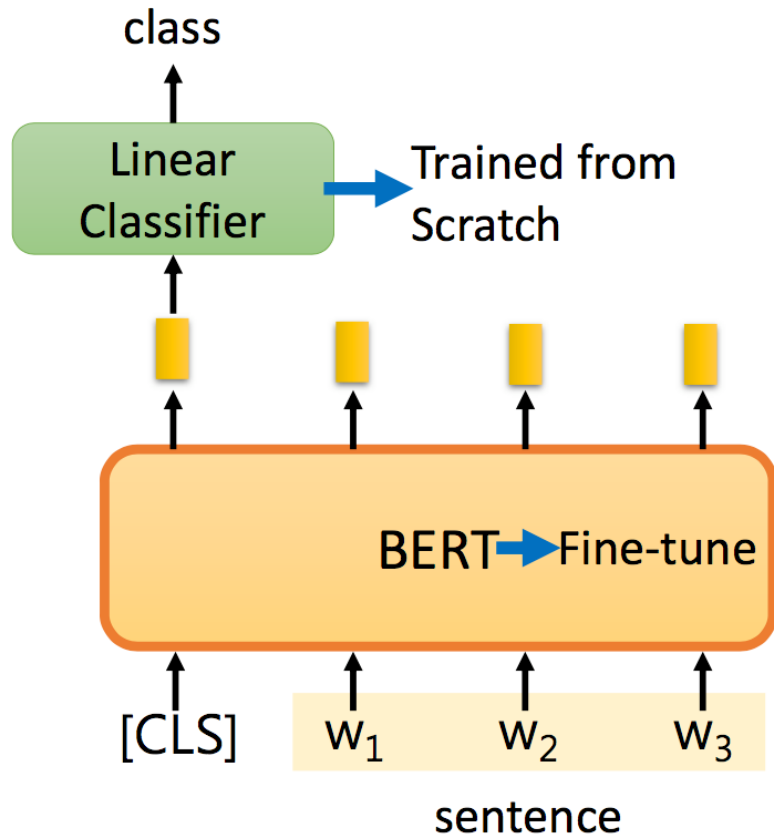
Bert训练方法Next Sentence Prediction: 预测与下一个句子的关系;

Approach 2: Next Sentence Prediction



□ BERT (Bidirectional Encoder Representations from Transformers)

Bert用途：文本分类；

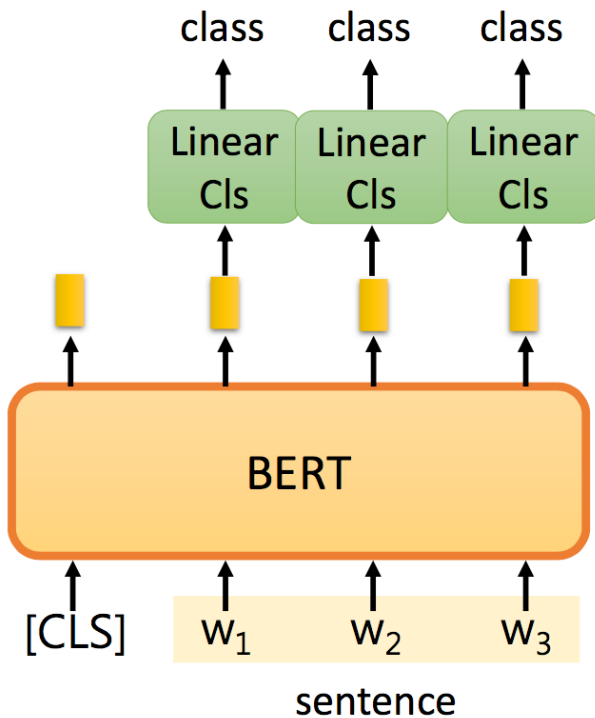


Input: single sentence,
output: class

Example:
Sentiment analysis (our
HW),
Document Classification

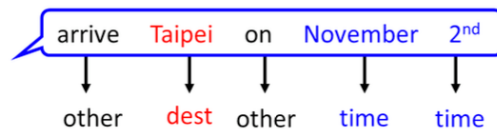
□ BERT (Bidirectional Encoder Representations from Transformers)

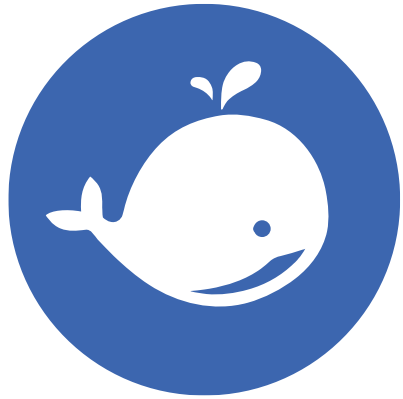
Bert用途：词性标注；



Input: single sentence,
output: class of each word

Example: Slot filling





**Thank
you**