

实体识别任务实践(三)-Bert

BERT简介



论文题目: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

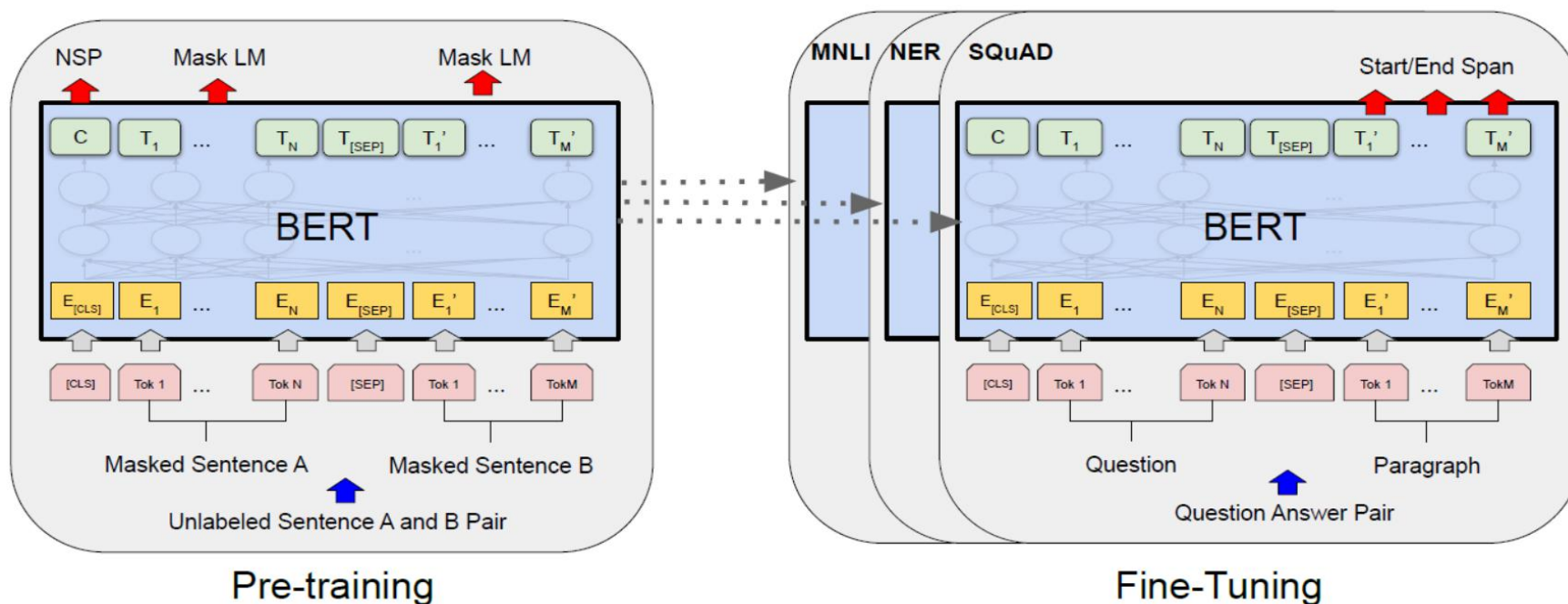
论文地址: <https://arxiv.org/abs/1810.04805>

论文源码: <https://github.com/google-research/bert>

BERT(Bidirectional Encoder Representation from Transformers)是2018年10月由Google AI研究院提出的一种预训练模型, 该模型在机器阅读理解顶级水平测试SQuAD1.1中表现出惊人的成绩: 全部两个衡量指标上全面超越人类, 并且在11种不同NLP测试中创出SOTA表现, 包括将GLUE基准推高至80.4% (绝对改进7.6%), MultiNLI准确度达到86.7% (绝对改进5.6%), 成为NLP发展史上的里程碑式的模型成就。

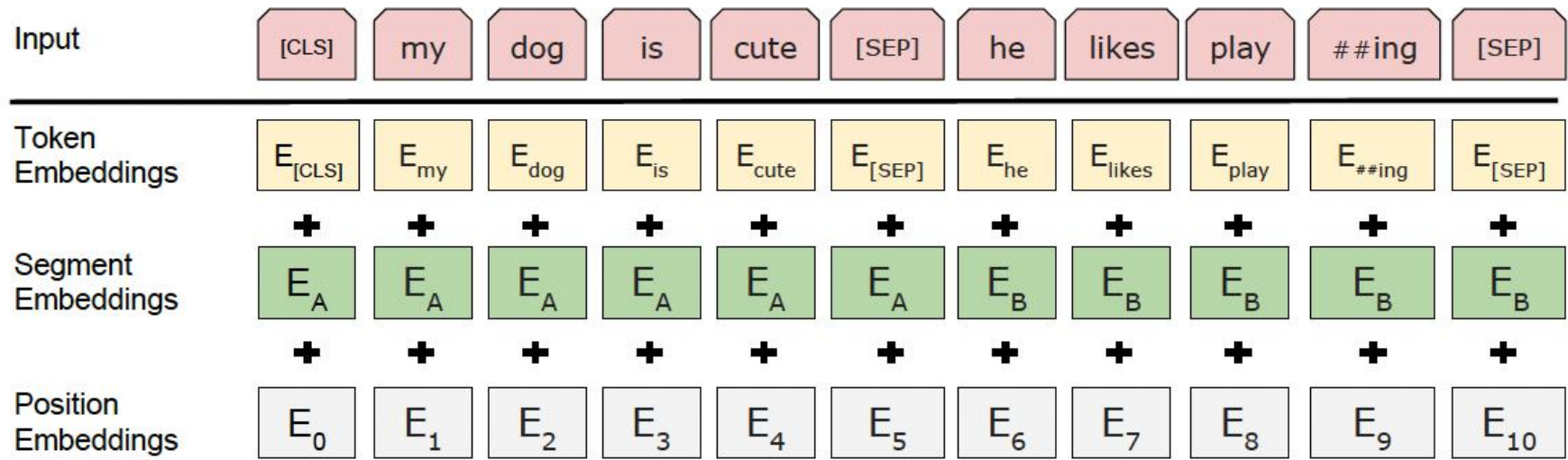
BERT框架

BERT整体框架包含pre-train和fine-tune两个阶段。pre-train阶段模型是在无标注的标签数据上进行训练，fine-tune阶段，BERT模型首先是被pre-train模型参数初始化，然后所有的参数会用下游的有标注的数据进行训练。



BERT是用了Transformer的encoder侧的网络，encoder中的Self-attention机制在编码一个token的时候同时利用了其上下文的token，其中‘同时利用上下文’即为双向的体现，而并非像Bi-LSTM那样把句子倒序输入一遍。

BERT的Embedding



- Token Embeddings是词向量，第一个单词是CLS标志，可以用于之后的分类任务
- Segment Embeddings用来区别两种句子，因为预训练不光做LM还要做以两个句子为输入的分类任务
- Position Embeddings和之前文章中的Transformer不一样，不是三角函数而是学习出来的

其中[CLS]表示该特征用于分类模型，对非分类模型，该符号可以省去。[SEP]表示分句符号，用于断开输入语料中的两个句子。

The Illustrated Transformer



Jay Alammam

Visualizing machine learning one concept at a time.

@JayAlammam on Twitter. [YouTube Channel](#)

[Blog](#) [About](#)

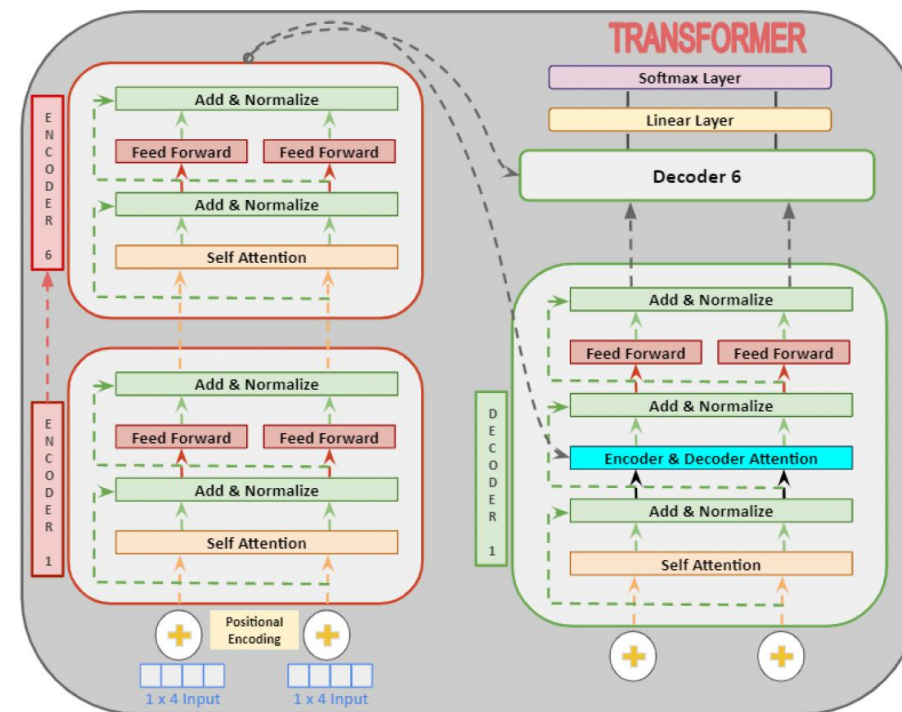
The Illustrated Transformer

Discussions: [Hacker News](#) (65 points, 4 comments), [Reddit r/MachineLearning](#) (29 points, 3 comments)

Translations: [Chinese \(Simplified\)](#), [French 1](#), [French 2](#), [Japanese](#), [Korean](#), [Russian](#), [Spanish](#), [Vietnamese](#)

Watch: MIT's [Deep Learning State of the Art](#) lecture referencing this post

- <https://jalammar.github.io/illustrated-transformer/>
- https://paddlepedia.readthedocs.io/en/latest/tutorials/pretrain_model/bert.html



Bert For NER

