

BERT :

标签: Google AI

2018

NLP 领域

机器阅读理解 · 善解

里程碑式 ( GLUE = 80.4% )

MultNLI : 86.7% ( ↑ 5.6% )

预训练语言模型

Model 特点:

① 深而窄  $\Rightarrow$  CV field's intuition  
12 层 中间层: 1024, 相比 transformer-2018

② Masked Language Model  
遮挡

③ Fine-tune-based model

Transformer-encoder-based model

Key words: ① Pre-training  $\Rightarrow$

免去了许多工程任务 需要针对特定任务修改体系架构的“需求”。BERT是第一个基于不微调的表示模型 (即应用时只需微调模型的几层即可)

Model to be pre-trained  $\Rightarrow$  Deep and bidirectional Model.

fine-tune based

# △ BERT 关键创新.

具体做法：(i) Mask

对 15 篇 大章，每篇随机遮盖 15%  
词汇。通过全向预测这些 masked words，作为  
attention，来对 model 进行初步 pre-training  
(ii) 下一句预测

从文章中挑选 20 万对语句 (总：40 万条)

其中 10 万对语句，为连续的两条上下文语句

10 万对语句，为不连续的两条上下文语句

通过 Transformer model 来识别这些 (20 万对) 语句，哪  
对是连续的，哪对是不连续的。

来对 model 进行进一步 pre-training

How to solve:

Deep bi-directional. { ② Deep

上下文全向预测 }

e.g.: I have a [MASK] and  
a banana.

③ Bidirectional (⇒ (i) 从后往前预测表征(mask)，为单向预测 ⇒ )

来自于 transformer  
文章

缺点：不能完整理解整个语句语义

(ii) 从后往前 + 从前往后 = 双向预测

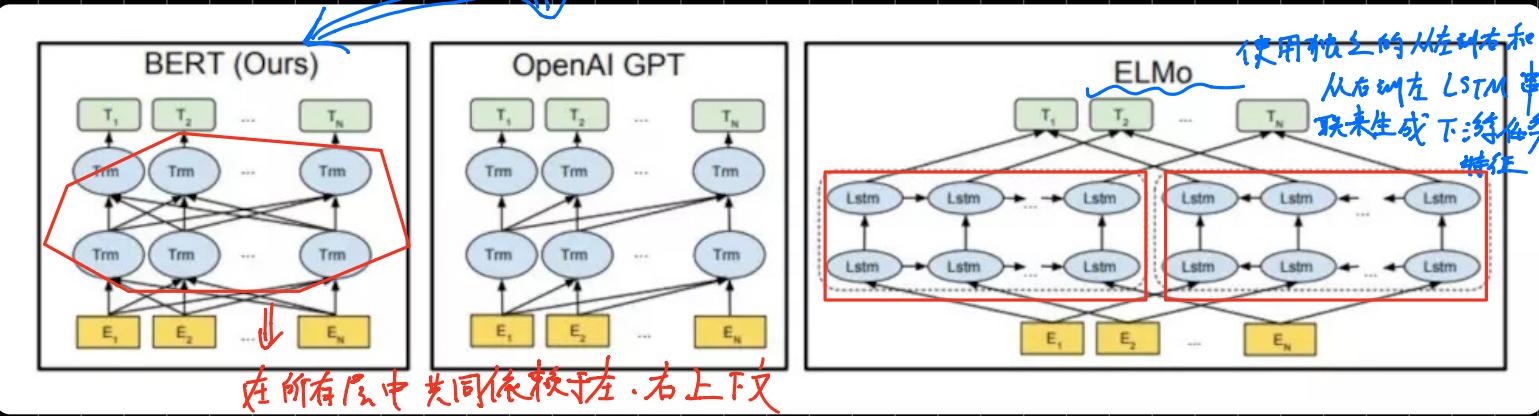
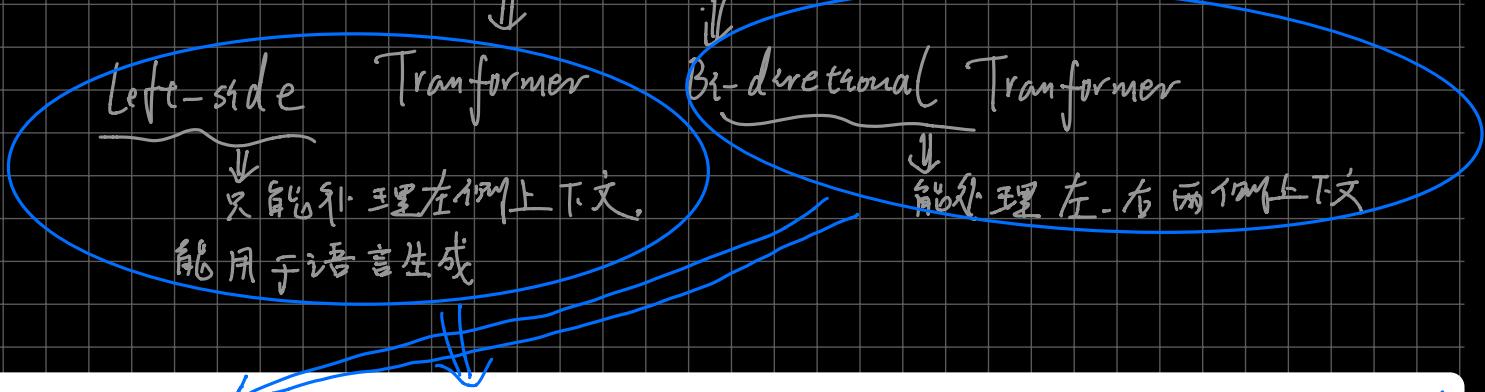
缺点：不能完整理解整个语句语义

④ Transformer ⇒ 来自 << Attention Is All You Need >>

Transformer 核心 ( attention 机制 )  
聚值

目的：对于一个 sentence，可启用多个 attention point，而不  
必局限于 单向 / 双向，可进行全向的预测  
序列串行处理

Transformer Decoder . Encoder :



## ⑤ Language Understanding

解决问题：

读《A Neural Probabilistic Language Model》

论文启发

神经概率语言模型：① 能运用  
word vector 来表达自然语言词汇语义

数值向量

② 如何给每个词汇，找到合适  
的 word vector.

观点：每一篇文章，天生是训练数  
据，而无需人工去标注产生

本质：Language Generative Model  $\Rightarrow$

语言生成模型。

相较于 Language Representation Model (语言表征模型) 而言 (BERT 属  
为此类)

Input representation (输入表征)： $\Rightarrow$  New embedding =

Token Embeddings + Position Embeddings + Segmentation Embeddings

减少 self-attention 机制忽略  
word position information 影响

表示两个 sentence 的  
分隔信息

# BERT 模型的影响：

① Deep Learning is representation Learning

⇒ 有了 BERT 模型的基础（有了 representation）  
各种不同的应用（specific task）都可以在  
BERT base 上进行微调就产生不错的效果

② Scale matters

⇒ BERT 模型得益于：

① 超大模型

（12 layer \* 1024 (normal)

24 layer \* 2048 (large)

）

② 超大数据（Wiki 2.5G  
文章）

③ 超大规模算力（毕竟是  
Google）

NLP 领域衡量指标：

GLUE: General Language Understanding Evaluation

通用语言理解评估

Multigenre NLI: Multi-Genre Natural Language Inference

多类体自然语言推理