

AI 大模型开发工程师 之大模型核心之算法

讲师：李希沅

掌握大模型核心三要素



数据

算法

算力

大模型核心之算法

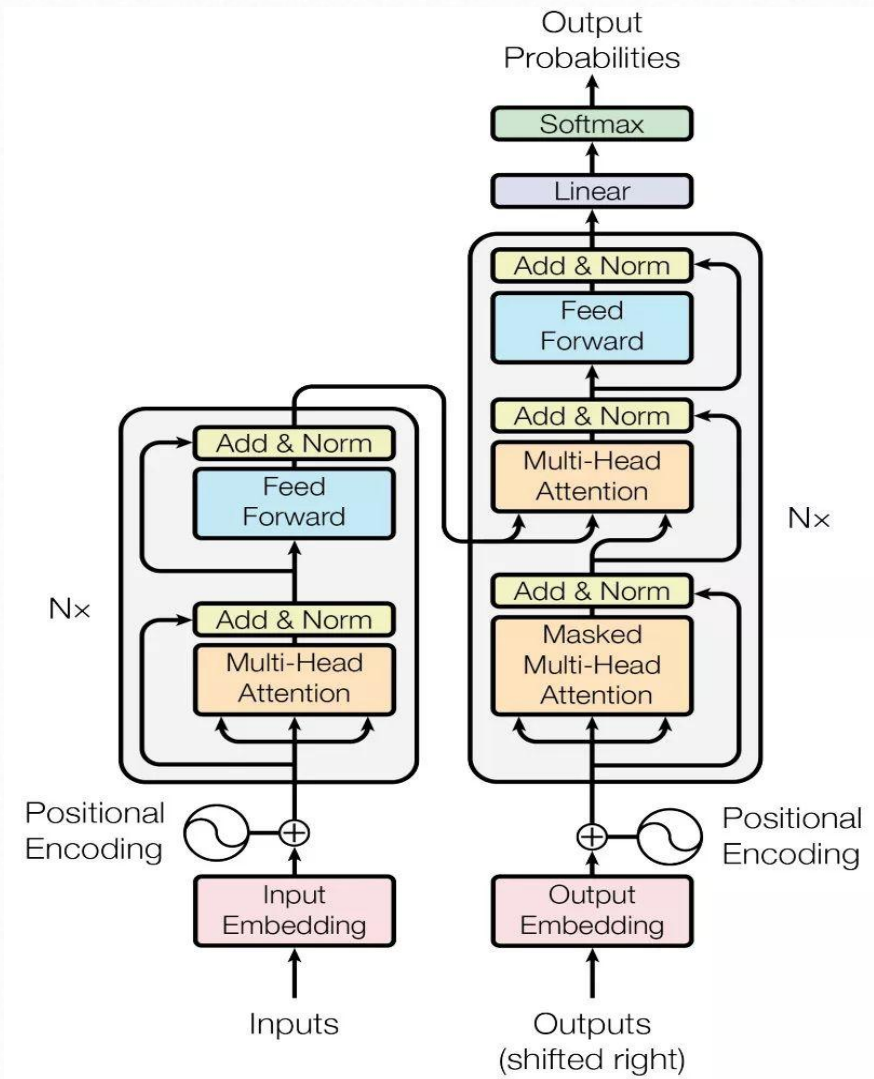


Figure 1: The Transformer - model architecture.

模型	结构	位置编码	激活函数	layer norm方法
原生 Transformer	Encoder-Decoder	Sinusoidal编码	ReLU	Post layer norm
BERT	Encoder	绝对位置编码	GeLU	Post layer norm
LLaMA	Casual decoder	RoPE	SwiGLU	Pre RMS Norm
ChatGLM-6B	Prefix decoder	RoPE	GeGLU	Post Deep Norm
Bloom	Casual decoder	ALiBi	GeLU	Pre Layer Norm

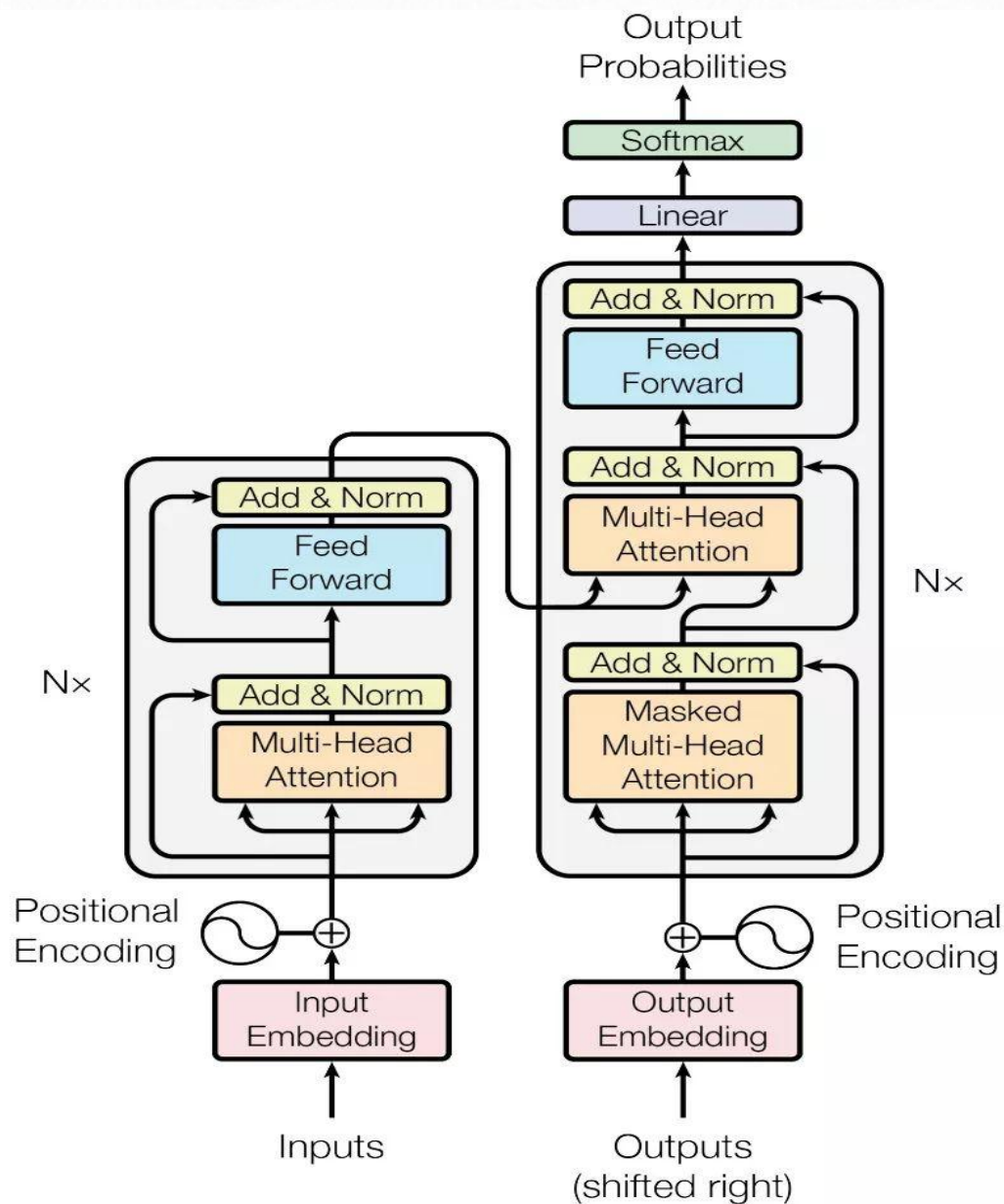
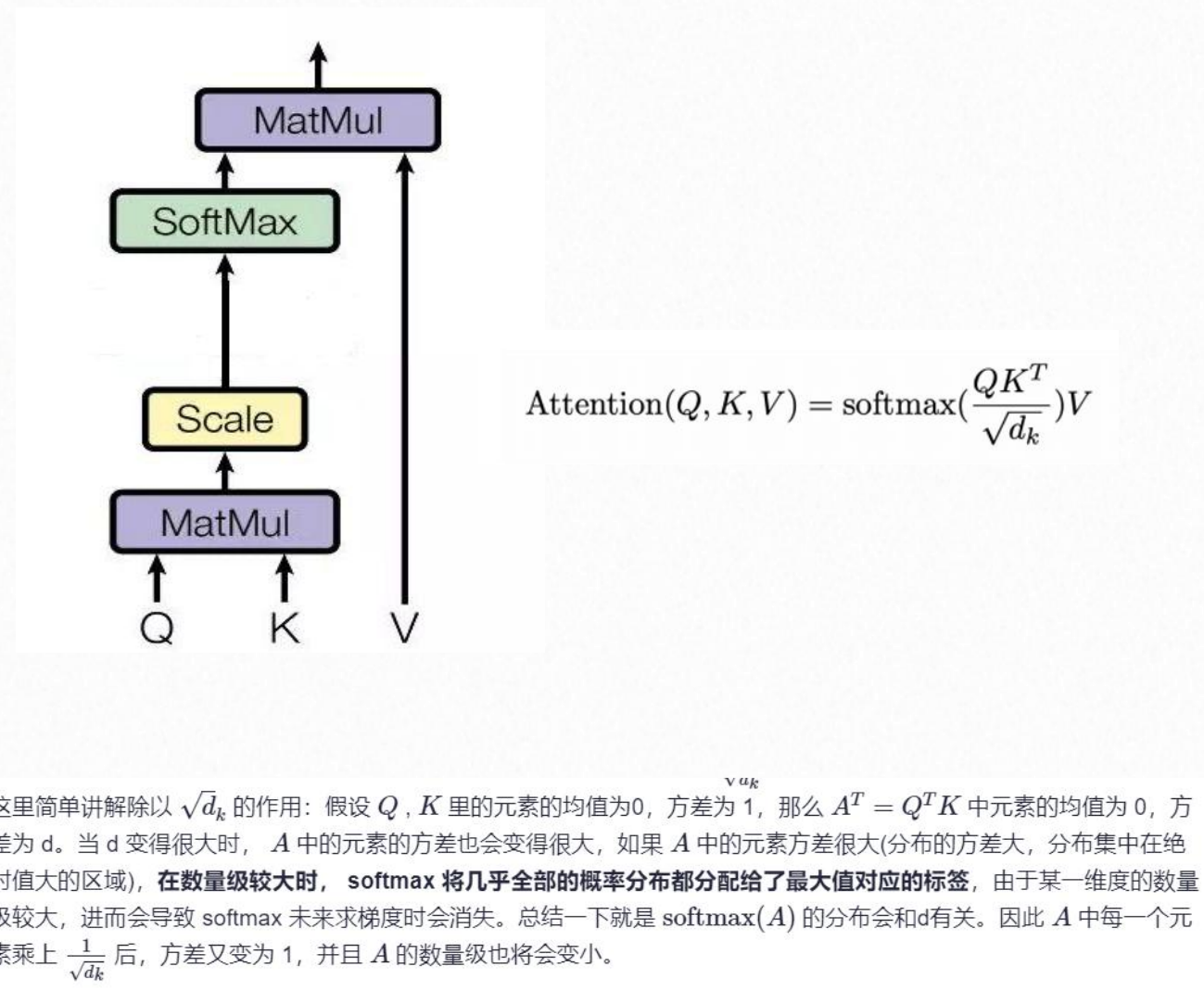


Figure 1: The Transformer - model architecture.



独热编码 (one-hot 编码)

	time	fruit	flies	like	a	an	arrow	ba
1 time	1	0	0	0	0	0	0	0
1 fruit	0	1	0	0	0	0	0	0
1 flies	0	0	1	0	0	0	0	0
1 like	0	0	0	1	0	0	0	0
1 a	0	0	0	0	1	0	0	0
1 an	0	0	0	0	0	1	0	0
1 arrow	0	0	0	0	0	0	1	0
1 banana	0	0	0	0	0	0	0	1

缺点1: 单词之间没关系

缺点2: 耗费存储

word2vec

Q 矩阵，对于任何一个独热编码的词向量都可以通过 Q 矩阵得到新的词向量

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = \begin{bmatrix} 10 & 12 & 19 \end{bmatrix} \quad \text{公式(8)}$$

优点1： 可以转换维度

优点2： 相似词之间的词向量之间也有了关系

有什么缺点？

ELMo 模型

有什么问题值得改进？

...very useful to protect **banks** or slopes from being washed away by river or rain...

...the location because it was high, about 100 feet above the bank of river...

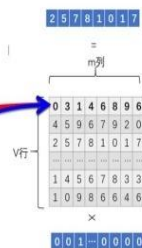
...The bank has plan to branch throughout the country...

...They throttled the watchman and robbed the bank...

(多义词) Bank:

1. 河岸

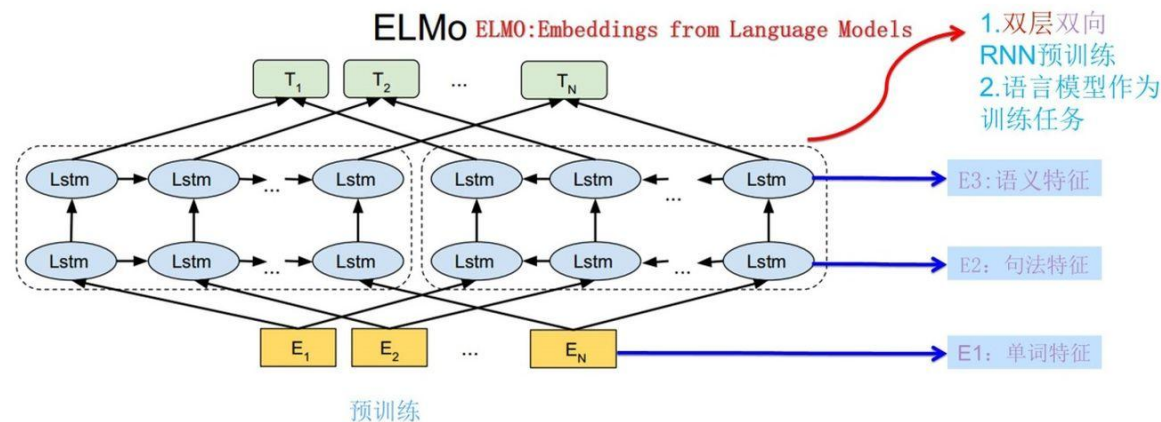
2. 银行



静态的Word Embedding!

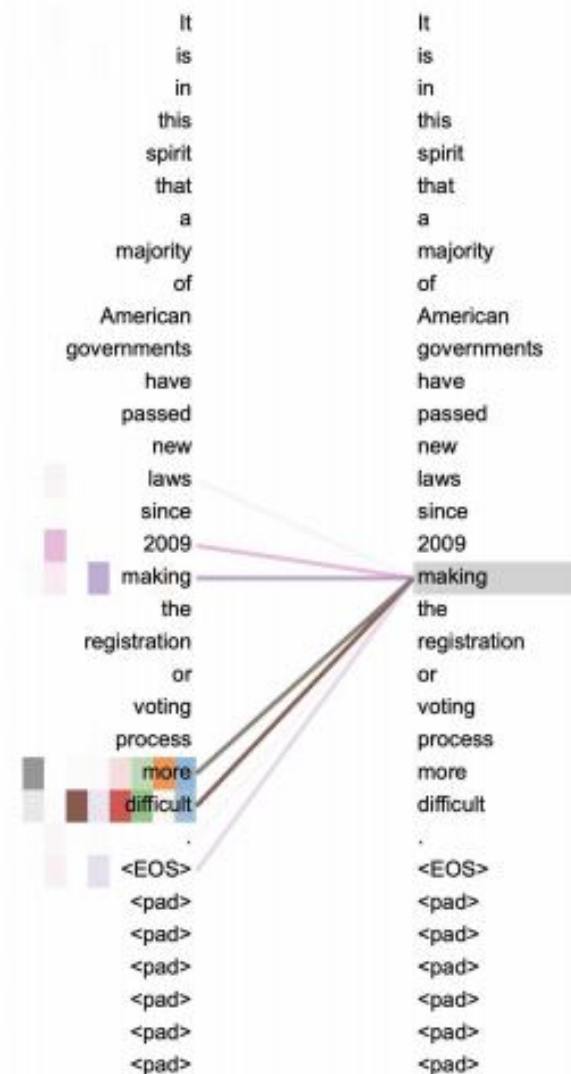
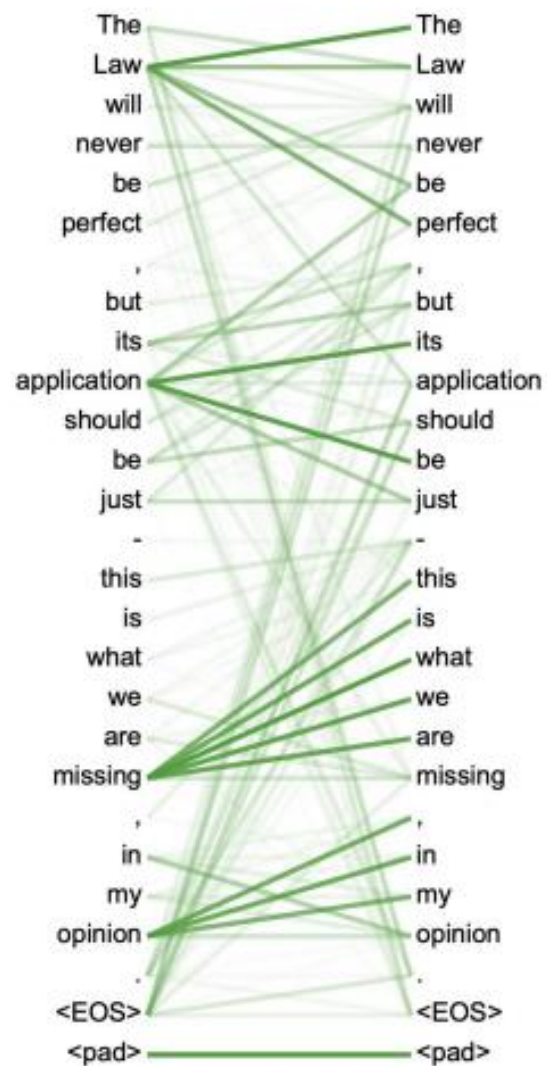


从WE到ELMO: 基于上下文的Embedding



NAACL 2018 最佳论文: Deep contextualized word representations

Multi-head Self Attention



谢谢观看