

# Transformer相关——（1）Encoder-Decoder 框架

📅 发表于 2021-08-15 | ⌚ 更新于 2021-08-20 | 📁 深度学习

| 📄 字数总计: 710 | ⌚ 阅读时长: 3分钟 | 👁 阅读量: 3463

## Transformer相关——（1）Encoder-Decoder 框架

### 引言

Transformer采用经典的encoder-decoder框架，是一个基于self-attention来计算输入和输出表示的模型，现已被应用于计算机视觉、自然语言处理等领域，都有非常好的效果。

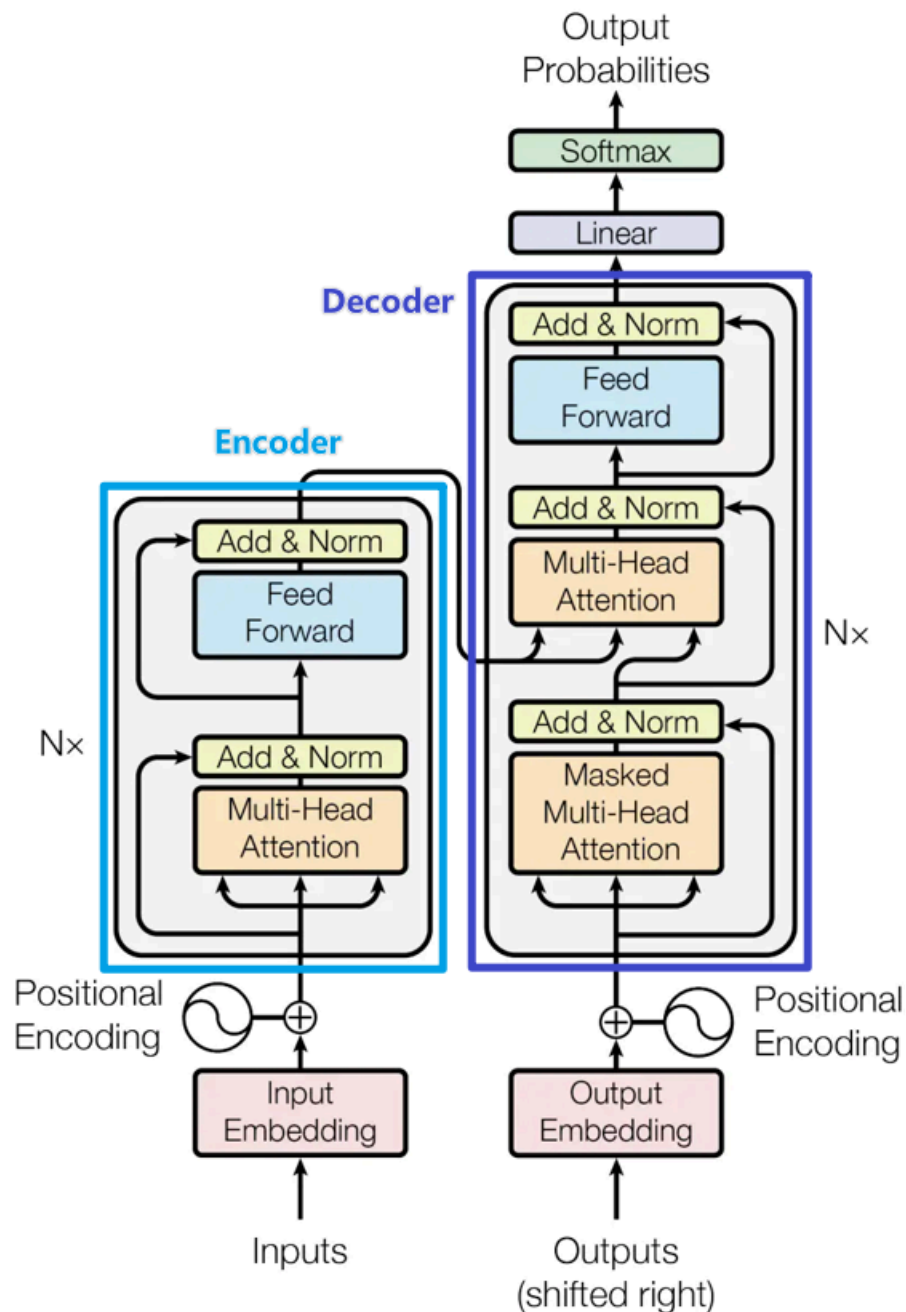


Figure 1: The Transformer - model architecture.

从这篇开始的“Transformer相关”博文，将尽可能覆盖Transformer触及的相关知识，一起手撕Transformer~

## Encoder-Decoder框架

Encoder-Decoder框架可以看作是一种深度学习领域的研究模式，即：将现实问题转化为一类可优化或者可求解的数学问题，利用相应的算法来实现这一数学问题的求解，然后再应用到现实问题中，从而解决了现实问题。

Encoder-Decoder框架分为Encoder（编码器）和Decoder（解码器）两个部分，分别实现 现实问题→数学问题的转化 和 求解数学问题→现实问题解决方案的转

化。

对于文本领域的Encoder-Decoder框架的实际模型案例比如，输入一个句子序列 $X_1, X_2, X_3 \dots$ 经过encoder进行非线性编码，获得一个向量 $C$ （中间语义表示），decoder根据这个向量和之前生成的历史信息去生成另外一个句子 $Y_1, Y_2, Y_3 \dots$ 。需要注意的是，这里的 $Y_i$ 除了受向量 $C$ 的影响，通常还受前序逐步生成的历史信息影响，即： $Y_i = \text{Decoder}(C, y_1, y_2, \dots, y_{i-1})$ 。

### Encoder：将现实问题转化为数学问题



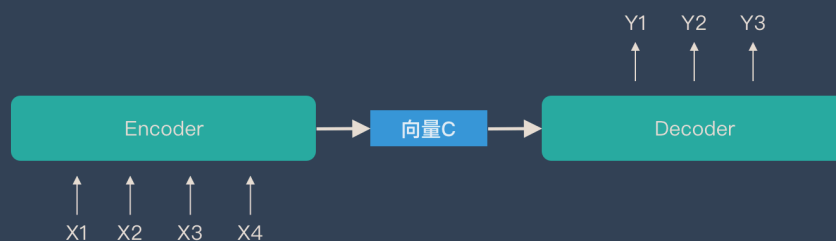
easyai

### Decoder：求解数学问题，并转化为现实世界的解决方案

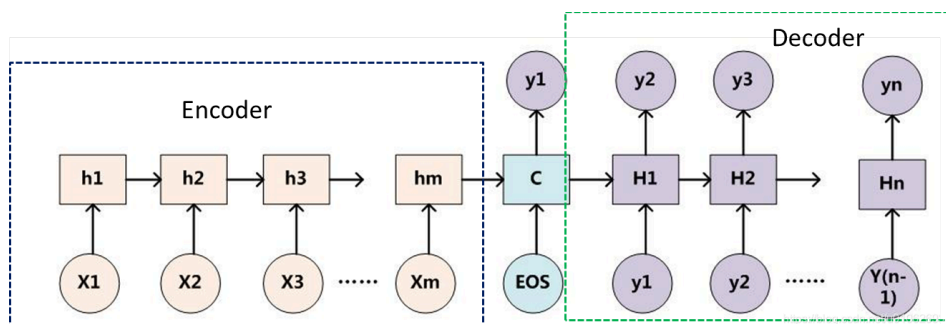


easyai

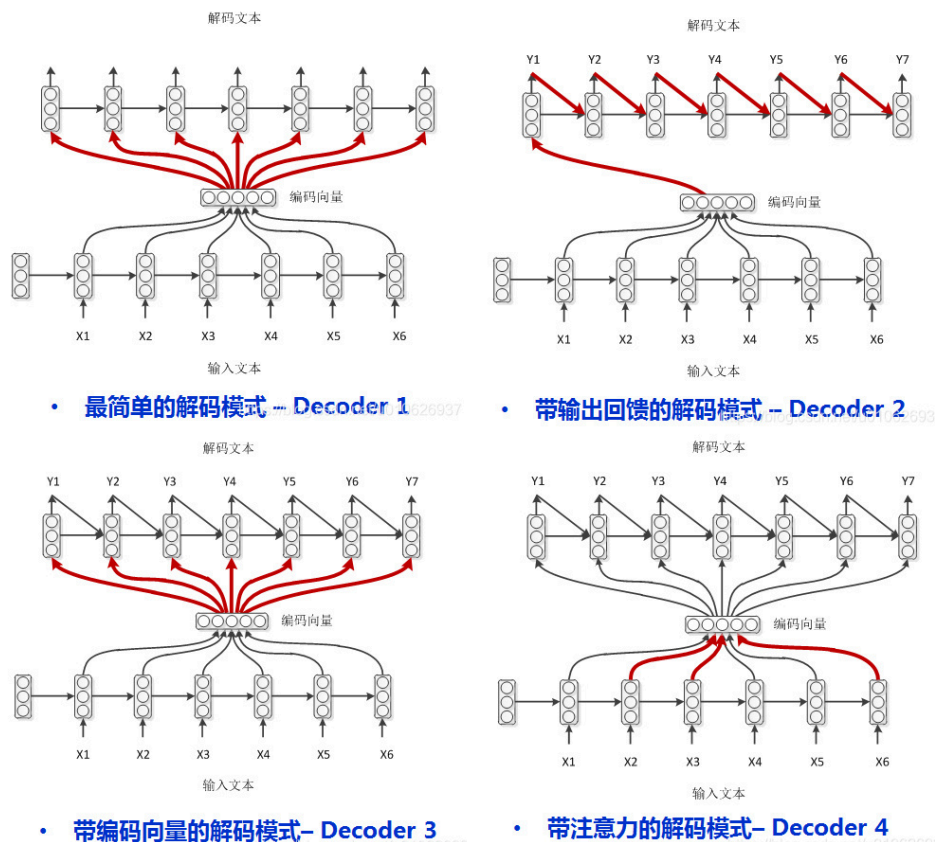
### 图解 Encoder-Decoder



easyai



## 四种Encoder-Decoder模式



## Encoder-Decoder框架特点

- 1 它是一个end-to-end的学习算法；
- 2 不论输入和输出的长度是什么，中间的向量 $C$ 的长度都是固定的（导致存在信息缺失问题）；
- 3 根据不同的任务可以选择不同的编码器和解码器（可以是CNN、RNN、LSTM、GRU等）。

## Encoder-Decoder框架应用

基于Encoder-Decoder框架设计的模型可以应用于：机器翻译、对话机器人、诗词生成、代码补全、文章摘要、语音识别、图像描述生成等方面。



## Encoder-Decoder框架缺点

基础Encoder-Decoder框架存在的最大问题在于信息缺失。

Encoder将输入（Source）编码为固定大小的向量的过程是一个“信息有损的压缩过程”，信息量越大，转化得到的固定向量中信息的损失就越大，这就得Decoder无法直接无关注输入信息的更多细节。输入的序列过长，先输入的内容携带的信息可能会被后输入的信息稀释掉或被覆盖了，那么解码的时候一开始就没有获得输入序列足够的信息，可能会导致模型效果比较差。

## 参考文献

Encoder-Decoder综述理解(推荐)

从Encoder-Decoder(Seq2Seq)理解Attention的本质

Encoder-Decoder 和 Seq2Seq