

## NLP中的Encoder-Decoder模型笔记

笔记本： 深度学习

创建时间： 2018/10/7 11:18

更新时间： 2018/10/7 14:37

作者： beyourselfwb@163.com

---

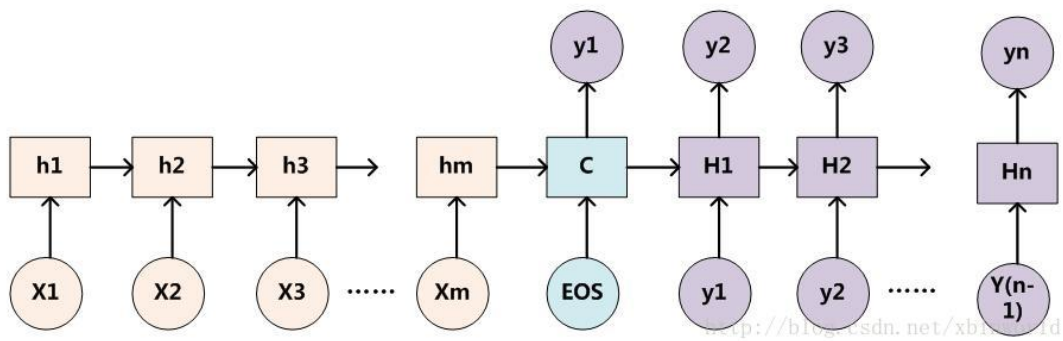
本文来自 大饼博士X 的CSDN 博客，全文地址请点击：

[https://blog.csdn.net/xbinworld/article/details/54605408?utm\\_source=copy](https://blog.csdn.net/xbinworld/article/details/54605408?utm_source=copy)

基本的Encoder-Decoder模型非常经典，但是也有局限性。最大的局限性就在于编码和解码之间的唯一联系就是一个固定长度的语义向量 $c$ 。也就是说，编码器要将整个序列的信息压缩进一个固定长度的向量中去。但是这样做有两个弊端，一是语义向量无法完全表示整个序列的信息，还有就是先输入的内容携带的信息会被后输入的信息稀释掉，或者说，被覆盖了。输入序列越长，这个现象就越严重。这就使得在解码的时候一开始就没有获得输入序列足够的信息，那么解码的准确度自然也就要打个折扣了[3]。

基本的Encoder-Decoder中Encoder和Decoder的唯一联系被压缩到一个固定长度的语义向量 $C$ ，这样做的局限性有：

- 1、 $C$ 无法完全表示整个序列的信息；
- 2、序列前面的内容会被后面的内容稀释，序列越长，稀释越明显。



基于上述局限性提出了 Attention Model

深度学习方法（九）：自然语言处理中的Attention Model注意力模型

链接：

<https://blog.csdn.net/xbinworld/article/details/54607525>

