总结循环神经网络处理语音识别应用的过程

循环神经网络RNN

RNN是一类用于处理序列数据的神经网络，在深度学习方面，图像识别、语音识别主要运用卷积神经网络（CNN），而文字语言处理主要运用循环神经网络（RNN）。

   首先是语言模型和序列生成。什么是语言模型呢？

语言模型

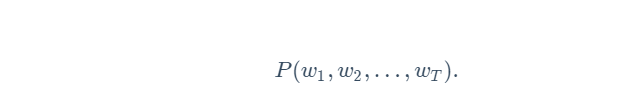
语言模型（language model）是自然语言处理的重要技术，可用于提升语音识别和机器翻译的性能。自然语言处理中最常见的数据是文本数据。

例如，在语音识别中，给定一段“厨房里食油用完了”的语音，有可能会输出“厨房里食油用完了”和“厨房里石油用完了”这两个读音完全一样的文本序列。如果语言模型判断出前者的概率大于后者的概率，我们就可以根据相同读音的语音输出“厨房里食油用完了”的文本序列。

假设一段长度为TT的文本中的词依次为w1,w2,…,wt。 ​

那么在离散的时间序列中，wt可看作在时间步(time step)的输出或标签。给定一个长度为T的词的序列w1,w2,…,wt。

语言模型将计算该序列的概率：

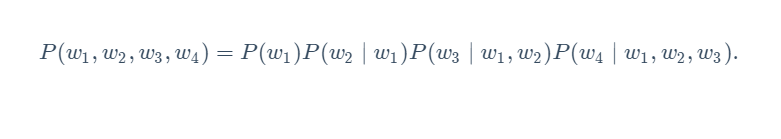


亦即词w1,w2,…,wt组合输出一个文本的概率

语言模型的计算

例如，假设序列w1,w2,…,wt​

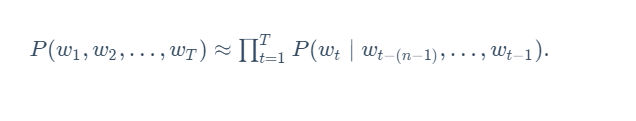
中的每个次是一次生成的，则一段含有4个词的文本序列的概率为：



n元语法

当序列长度增加时，计算和存储多个词共同出现的概率的复杂度会呈指数级增加。n 元语法通过马尔可夫假设（虽然并不一定成立）简化了语言模型的计算。这里的马尔可夫假设是指一个词的出现只与前面n个词相关，即n 阶马尔可夫链（Markov chain of order n ）。

如果基于n−1 n−1n−1阶马尔可夫链，我们可以将语言模型改写为



如何得到一个语言模型呢？  --首先需要一个语料库（可以是中文或者英文），将这些语料组成一个字典，将每个单词转化为字典索引，在每一个句子末尾加上EOS（分割标志），如果在新的句子中出现语料库中没有的单词，应该将这个陌生的单词替换为“UNK”（未知词的标志）然后根据UNK建立一个概率模型。

当训练过序列模型后，如果想要知道到底学到了什么，一种非正式的方法就是对序列模型进行下采样。就是说首先初始化，经过激活函数（常用sigmoid函数）后，得到第一个输出，对该输出进行下采样，并将下采样的结果（如the）作为第二层的输出，并经过激活后得到第二个输出直到达到所设置的时间步。

但是在该过程中会出现一些未知的标识，如果你需要确保算法不输出这些标识，那么就需要在采样的过程中拒绝产生的未知标识。如果采样得到未知标识的单词，可以忽略不计该单词，对输出进行重采样，直到输出的是语料库的单词。如果并不介意未知标识的单词，就不需要这样做。