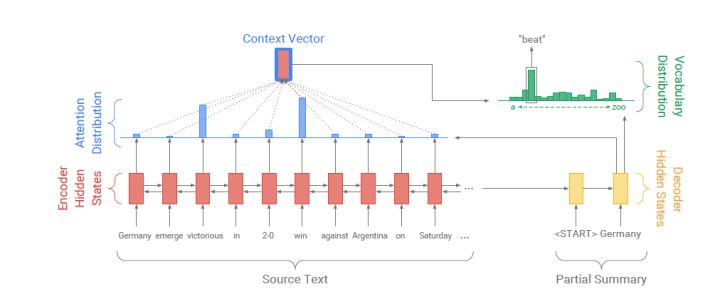
传统带有注意力机制的seq2seq模型输出的是针对输出词汇表的一个概率分布，而Pointer Networks输出的则是针对输入文本序列的概率分布。

因为输出元素来自输入元素的特点，Pointer Networks特别适合用来直接复制输入序列中的某些元素给输出序列。而事实证明，后来的许多文章也确实是以这种方式使用Pointer Networks的。

传统seq2seq模型正如我们开始描述的那样，通过Attention Mechanism将encoder的隐状态和decoder的隐状态结合成一个中间向量C，然后使用decoder解码并预测，最后经由softmax层得到了针对词汇表的概率分布，从中选取概率最高的作为当前预测结果。



而Pointer-Generator模型除了上述过程，还加入了Pointer Networks的部分。就是用它来复制源文本中的单词。简单来说，在每一次预测的时候，通过传统seq2seq模型的预测（即softmax层的结果）可以得到针对词汇表的概率分布（图7中绿色柱形图），然后通过Pointer Networks可以得到针对输入序列的概率分布（图7中蓝色柱形图），对二者做并集就可以得到结合了输入文本中词汇和预测词汇表的一个概率分布，这样一来模型就有可能直接从输入文本中复制一些词到输出结果中。当然，直接这样操作未必会有好的结果，因此作者又加入了一个Pgen来作为软选择的概率。Pgen的作用可以这样理解：决定当前预测是直接从源文本中复制一个词过来还是从词汇表中生成一个词出来。

https://pic4.zhimg.com/80/v2-4d497d0d859cfd528aa1786b7e6857fb_hd.jpg

https://pic4.zhimg.com/80/v2-51a651f6965a08e140fc816e1083fd83_hd.jpg

