Attention for Seq2Seq model

2021年2月4日 14:24

1. Attention for Sequence-to-Sequence

- 链接：https://www.youtube.com/watch?v=XhWdv7ghmQQ&t=228s

Another related blogs: https://lena-voita.github.io/nlp_course/seq2seq_and_attention.html

- Shortage for seq2seq

  如果输入的句子很长的话，seq2seq模型会记不住整个句子

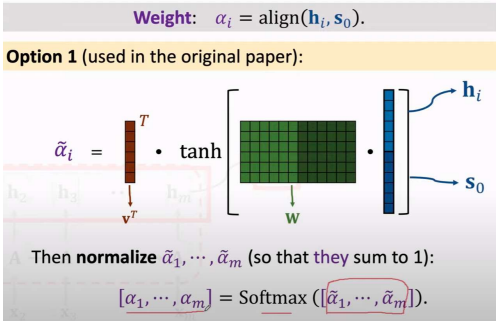
  Seq2seq模型做机器翻译，如果没有加attention，那么当word大于20的话BLUE分数会下降，而加了attention就会避免这个问题

  解决seq2seq遗忘问题最有效的方法是attention机制！！！

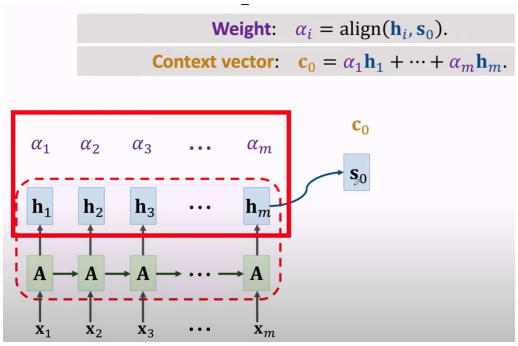
  attention使得seq2seq模型不要忘记source input， 并且会让decoder知道哪个地方应该被focus

  attention可以极大的增加计算的准确率，但是计算量相当的大，没加attention之前，时间复杂度O(m+t)， 加了之后的时间复杂度O(m·t),其中m为权重的个数，表示输入序列的长度，t为states的个数，表示输出序列的长度

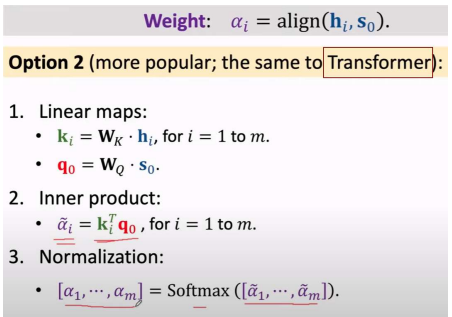
  1. 每个状态对应一个权重参数
  2. 有两种方法来计算权重参数：权重参数为decoder的状态向量和encoder的隐藏向量之间的相关性
     - 方法1：v^T 和 W是可训练的！！！



- seq2seq+attention 计算过程



方法2： 主流的，transformer用的也是这个



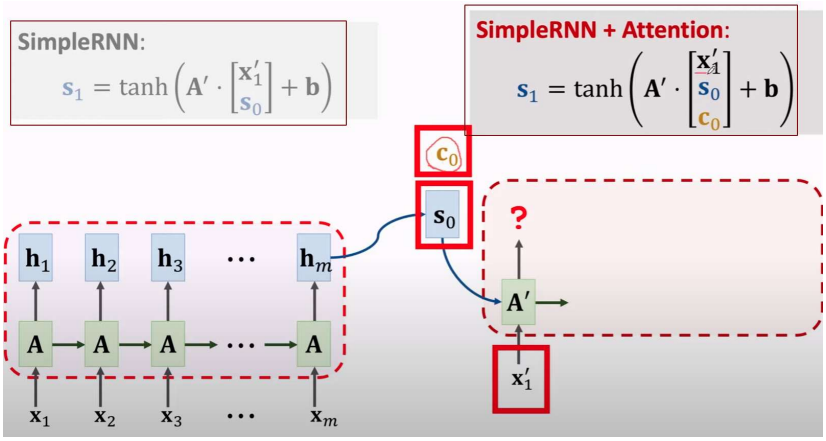
attention表现在每次在decoder的时候生成的状态向量s将会和encoder中的所有状态向量h做一个align

- Decoder 状态更新

  1. c_0为加权平均
  2. 更新s_1使用三个向量的聚合以及线性变换

- Decoder 状态更新

1. c_0为加权平均
2. 更新s_1使用三个向量的聚合以及线性变换
3. 下一步计算c_1的时候，需要重新计算权重参数



- attention weight visualization

1. 线条越粗说明目标之间的相关性越大！