

第八天

2019年6月17日 9:32

seq2seq的优化

1. teacher forcing机制:

- a. 在seq2seq中避免一步错，步步错的情况
- b. 把真实值作为下一步的输入，能够加快模型的训练速度
- c. 在输出确定的情况下：可以在循环中使用teacher forcing
- d. 如果输出不确定，应该在循环外使用

2. Attention

- a. 初始化decoder的隐藏状态 Z_i
- b. 和encoder每个时间步进行计算，得到的结果进行softmax，转化为概率
- c. 概率和encoder每个时间步记性相乘相加，得到 C_i
- d. C_i 作为decoder的输入，得到输出

3. attention的计算:

- a. 计算attention weight
- b. context vector
- c. attention result

4. bahdanau 和Luong attention的区别

- a. 计算attention结果的位置不同
 - i. baha 在decoder的每个时间步之前计算attention的结果，bahd使用的双向GRU计算attention的结果
 - ii. Luong 在decoder每个时间步之后计算attention的二级果，使用单向多层GRU
- b. attention weight的区别
 - i. bahd: $V * \tanh(Wz_i - I + Uh_j)$
 - ii. luong:
 - 1) general :进行矩阵变换后进行矩阵乘法
 - 2) dot: 对应位置相乘
 - 3) concat: