

# Homework

## Homework

### Part 1

#### Process

(g)

(f)

### Part 2

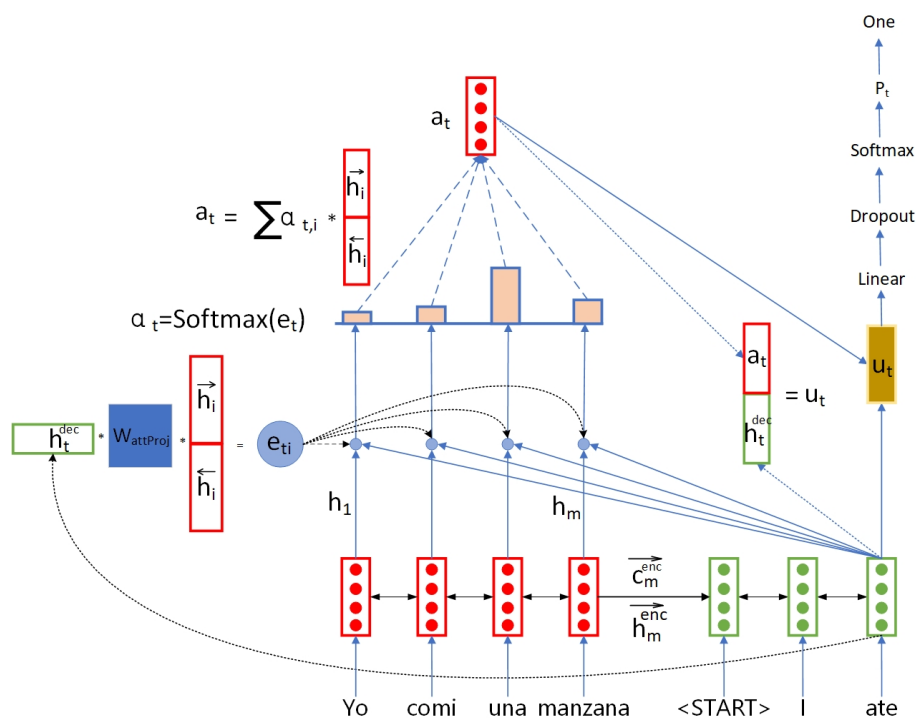
(a)

(b)

(c)

## Part 1

### Process



(g)

- 使用 masks 将句子中的 pad token 的分数赋值为  $-\inf$ ，从而使得 softmax 作用后获得的 attention 分布中，pad token 的 attention 概率值近似为 0
- attention score / distributions 计算的是 decoder 中某一时间步上的 target word 对 encoder 中的所有 source word 的注意力概率，而 pad token 只是用于 mini-batch，并没有任何语言意义，target word 无须为其分散注意力，所以需要使用 masks 过滤掉 pad token

(f)

	优点	缺点
点 积 注 意 力	不需要额外的线性映射层	$s_t, h_t$ 必须有同样的纬度，且效果可能较差
乘 法 注 意 力	$s_t, h_t$ 不需要有同样的纬度并且因为可以使用高效率的矩阵乘法，比加法注意力要更快更省内存	增加了训练参数
加 法 注 意 力	高维时的表现更好	训练参数更多（两个参数矩阵以及注意力的纬度），比较慢

Part 2

(a)

1. Identify the error in the NMT translation.
2. Provide a reason why the model may have made the error (either due to a specific linguistic construct or specific model limitations).
3. Describe one possible way we might alter the NMT system to fix the observed error.

Below are the translations that you should analyze as described above. Note that out-of-vocabulary

**Source Sentence:** *Aquí otro de mis favoritos, “La noche estrellada”.*

**Reference Translation:** *So another one of my favorites, “The Starry Night”.*

**NMT Translation:** *Here’s another favorite of my favorites, “The Starry Night”.*

- Error: favorite of favorite
- Reason: Specific linguistic construct. This word pair rarely appears.
- Solution: Try more corpus.

**Source Sentence:** *Ustedes saben que lo que yo hago es escribir para los niños, y, de hecho, probablemente soy el autor para niños, más leído en los EEUU.*

**Reference Translation:** *You know, what I do is write for children, and I'm probably America's most widely read children's author, in fact.*

**NMT Translation:** *You know what I do is write for children, and in fact, I'm probably the author for children, more reading in the U.S.*

- Error: more in the U.S.
  - Reason: Specific linguistic construct. False meaning may cause by the misunderstanding by encoder or wrong expression by decoder.  
(模型解释能力不足)
  - Solution: Add more numbers of the hidden layer.
- 

**Source Sentence:** *Un amigo me hizo eso – Richard Bolingbroke.*

**Reference Translation:** *A friend of mine did that – Richard Bolingbroke.*

**NMT Translation:** *A friend of mine did that – Richard*

- Error: Bolingbroke.
  - Reason: Model limitations. In the corpus, there is no such word.
  - Solution: Add this kind of name to the model.
- 

**Source Sentence:** *Solo tienes que dar vuelta a la manzana para verlo como una epifanía.*

**Reference Translation:** *You've just got to go around the block to see it as an epiphany.*

**NMT Translation:** *You just have to go back to the apple to see it as a epiphany.*

- Error: go back to the apple
  - Reason: Model limitations. "manzana" has multiple meanings while in Spanish, "block" appears more often than "apple". However, in the corpus, meaning "appear" may appear more often or same as meaning of "block".
  - Solution: Add more sentences containing the "manzana" of meaning "block"
- 

**Source Sentence:** *Ella salvó mi vida al permitirme entrar al baño no de la sala de profesores.*

**Reference Translation:** *She saved my life by letting me go to the bathroom in the teachers' lounge.*

**NMT Translation:** *She saved my life by letting me go to the bathroom in the women's room.*

- Error: women
  - Reason: Model limitations. In the corpus, "women" appears more often than "teacher". As a result, translation may come from the bias of the training data
  - Solution: Add more sentences containing "teacher"
-

**Source Sentence:** *Eso es más de 100,000 hectáreas.*

**Reference Translation:** *That's more than 250 thousand acres.*

**NMT Translation:** *That's over 100,000 acres*

- Error: 100,000 acres
- Reason: Model limitations. 单位换算错误
- Solution: Add more sentences about acres and hectáreas.

(b)

pass

(c)

i.

**Source Sentence  $s$ :** *el amor todo lo puede*

**Reference Translation  $r_1$ :** *love can always find a way*

**Reference Translation  $r_2$ :** *love makes anything possible*

**NMT Translation  $c_1$ :** *the love can always do*

**NMT Translation  $c_2$ :** *love can make anything possible*

Compute the scores for  $c_1, c_2, \lambda_i = 0.5 \ i \in \{1, 2\}$

$c_1$ :

$$p_1 = \frac{0 + 1 + 1 + 1 + 0}{5} = 0.6$$

$$p_2 = \frac{0 + 1 + 1 + 0}{4} = 0.5$$

$$r^* = 4$$

$$BP = 1$$

$$BLEU_{c_1} = 1 \times \exp(0.5 \times \log(0.6) + 0.5 \times \log(0.5)) = 0.5477$$

$c_2$ :

$$p_1 = \frac{1 + 1 + 1 + 1 + 0}{5} = 0.8$$

$$p_2 = \frac{1 + 0 + 0 + 1}{4} = 0.5$$

$$r^* = 4$$

$$BP = 1$$

$$BLEU_{c_2} = 1 \times \exp(0.5 \times \log(0.8) + 0.5 \times \log(0.5)) = 0.632$$

According to the BLEU Scores,  $c_2$  is the better one.

And I think they are both great.

ii.

We lost Translation  $r_2$ , recompute.

Source Sentence  $s$ : *el amor todo lo puede*  
Reference Translation  $r_1$  : *love can always find a way*  
NMT Translation  $c_1$  : *the love can always do*  
NMT Translation  $c_2$  : *love can make anything possible*

Compute the scores for  $c_1, c_2, \lambda_i = 0.5 \ i \in \{1, 2\}$

$c_1$ :

$$\begin{aligned}p_1 &= \frac{0 + 1 + 1 + 1 + 0}{5} = 0.6 \\p_2 &= \frac{0 + 1 + 1 + 0}{4} = 0.5 \\r^* &= 6 \\BP &= \exp(-0.2) \\BLEU_{c_1} &= \exp(-0.2) \times \exp(0.5 \times \log(0.6) + 0.5 \times \log(0.5)) = 0.4484\end{aligned}$$

$c_2$ :

$$\begin{aligned}p_1 &= \frac{1 + 1 + 1 + 1 + 0}{5} = 0.4 \\p_2 &= \frac{1 + 0 + 0 + 0}{4} = 0.25 \\r^* &= 6 \\BP &= \exp(-0.2) \\BLEU_{c_2} &= \exp(-0.2) \times \exp(0.5 \times \log(0.4) + 0.5 \times \log(0.25)) = 0.2589\end{aligned}$$

According to the BLEU Scores,  $c_1$  is the better one.

And I think they are both great.

iii.

*Due to data availability, NMT systems are often evaluated with respect to only a single reference translation. Please explain (in a few sentences) why this may be problematic.*

We can see from above, many reference translations can make our translation's BLEU score close to its real expression. Because a sentence can have lots of correct translation with totally different word.

- 如果我们使用单一参考翻译，它增加了好翻译由于与单一参考翻译有较低的 **n-gram overlap**，而获得较差的BLEU分数的可能性。例如上例中，如果删去的参考翻译是  $r_1$ ，那么将使得  $c_1$  的BLEU分数变低。
- 如果我们增加更多的参考翻译，就会增加一个好翻译中 **n-gram overlap** 的几率，这样我们就有可能使好翻译获得相对较高的BLEU分数。

iv.

*List two advantages and two disadvantages of BLEU, compared to human evaluation, as an evaluation metric for Machine Translation.*

#### 优点

- 自动评价，比人工评价更快，方便，快速
- BLEU的使用普及率较高，方便模型之间的效果对比

#### 缺点

- 结果并不稳定，由于核心思想是 **n-gram overlap**，所以如果参考翻译不够丰富，会导致出现较好翻译获得较差BLEU分数的情况
- 不考虑语义与句法
- 不考虑词法，例如上例中的**make**和**makes**
- 未对同义词或相似表达进行优化