

1. (g). It eliminates the influence of padded sentences on the attention matrix. It is necessary because we don't want the padded sentences to have some unexpected changes on the attention matrix.

(j).

	dot product	multiply	additive
flexibility	low	median	high
efficiency	high	median	low

2. (a). i. error: Aquí otro de mis → Here's another favourite of my favourites

Reason: Maybe because there are several words in this sentence that means "favourite" in English.

Fix: Maybe add another matrix in the decoder LSTM to avoid two same words in a sentence.

ii). error: probablemente ... leido → "I'm probably ... more reading"

Reason: The model can't recognize the word ordering in Spanish.

Fix: Maybe delete the comma in the sentence and add another RNN to add back the comma after English translation.

iii). error: "Richard Bolingbroke" → "Richard <UNK>"

Reason: Can't recognize names

Fix: Add "name" in vocabulary and learn to "copy" from source text.

iv). Error: "dar vueltas ... manzana" → "have ... apple"

Reason: can't understand Spanish idioms

Fix: Have the machine trained on a well-labeled idioms dataset.

v). Error: "baño ... profesores" → "bathroom ... room"

vi). Error: "baño ... profesores" → "bathroom ... room"

Reason: No gender indication in the original sentence or teacher but shows women in translation.

Fix: Have the machine trained on a well-labeled profession dataset

vii). Error: = 100,000 hectares"  $\rightarrow$  = 10000 acres"

Reason: Wrong measurement transformation.

Fix: Check the condition if it contains measurements

If so, change the number into English measurements.

If not, directly go to the machine.

b).

$$c). i). C_1: P_1 = \frac{0+1+1+1+0}{5}$$

$$= \frac{3}{5}$$

$$P_2 = \frac{0+1+1+0}{4}$$

$$= \frac{1}{2}$$

$$r^* = 4$$

$$BP = 1$$

$$BLEU = 1 \cdot e^{(0.5 \cdot \log(\frac{3}{5}) + 0.5 \cdot \log(\frac{1}{2}))}$$

$$\approx 0.77$$

$$C_2: P_1 = \frac{1+1+0+1+1}{5}$$

$$= \frac{4}{5}$$

$$P_2 = \frac{1+1}{4}$$

$$BP = 1$$

$$BLEU = 1 \cdot e^{(0.5 \cdot \log(\frac{4}{5}) + 0.5 \cdot \log(\frac{2}{4}))}$$

$$\approx 0.82$$

According to BLEU,  $C_2$  is better. I agree, because the mean is the same to the reference sentence.

$$ii). C_1: P_1 = \frac{3}{5}$$

$$P_2 = \frac{1}{2}$$

$$1 + 2 = 2$$

$$BP=1$$

$$BLEU = 1 - e^{(0.5 \cdot \log(\frac{3}{5}) \cdot 0.5 + 0.5 \cdot \frac{1}{2})}$$

$$= 0.77$$

$$C_2: P_1 = \frac{2}{5}$$

$$P_2 = \frac{1}{4}$$

$$BLEU = 0.69$$

$C_1$  is better according to BLEU. However, I don't think it has the same meaning as the reference translation.

iii). It may be problematic because the meaning of a sentence can be conveyed in different ways. So, if we only have one reference sentence, we may give sentences that have same mean with different wording a very low score.

c). advantage: 1. It is quite simple

2. It is fast to calculate

disadvantage: 1. It can't judge the structure of sentence

2. It can't judge the meaning of sentence very completely.