

## 1. Neural Machine Translation with RNNs (45 points)

g) All padded word values are  $-\infty$ , so after softmax, the attention distributed to padded words are zero.

The paddings have no meaning, and the attention score to them must be as smaller as possible.

i) BLEU=22.66417215453647

j)

	advantage	disadvantage
$e_{t,i} = s_t^T h_i$	Simple, fast	s and h must have same dimension.
$e_{t,i} = s_t^T W h_i$	s and h don't need to have same dimension.	Add model parameters(W) to be trained, and model is more complex and takes longer to run.
$e_{t,i} = v^T (W_1 h_i + W_2 s_t)$	Perform better	Model is more complex, and the dimension of v is a hyperparameter.

## 2. Analyzing NMT Systems (30 points)

a)

- i. (2 points) **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:**

**Reason:** low-resource training language pairs.

**Possible Fix:** add more training data or train more epochs.

- ii. (2 points) **Source Sentence:** *Ustedes saben que lo que yo hago es escribir para los niños, y, de hecho, probablemente soy el autor para niños, ms ledo 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 reading in the U.S.' is of incorrect grammar.

**Reason:** Sentence is too long, and decoder of model can't generate correct result from a fixed-length vector encoded by encoder, even with attention mechanism.

**Possible Fix:** Increase hidden and cell dimension of LSTM.

- iii. (2 points) **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 <unk>*

**Error:** <unk>

**Reason:** Reference translation contains word which is out-of-vocabulary.

**Possible Fix:** Add training data, and increase number of words in the vocabulary.

- iv. (2 points) **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:** apple

**Reason:** 'manzana' has multiple meanings: apple, block, etc. In training dataset, 'manzana' means 'apple' in more cases.

**Possible Fix:** Add more training data on 'manzana' represents 'block'

- v. (2 points) **Source Sentence:** *Ella salvó mi vida al permitirme entrar al baño 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:** Too many language pairs about women in the dataset.

**Possible Fix:** Add language pairs about teacher in the dataset.

- vi. (2 points) **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

**Reason:** In training dataset, acres are directly modified by numbers, and hectareas are modified by "millions", "thousands", etc.

**Possible Fix:** add training data with numbers modifying hectareas directly.

b)

1) **source:** Lo haca en secreto,

**reference:** She did it in secret.

**NMT:** I was doing it in secret.

**error:** I

**reason:** the model hasn't learnt the ability to distinguish I and she.

**possible fix:** add more training language pairs containing 'she'

2) **source:** Es poco frecuente llegar realmente a presenciar un tmpano rodante.

**reference:** It's a very rare occasion that you get to actually witness an iceberg rolling.

**NMT:** It's very often going to be able to see an iceberg <unk>

**error:** often

**reason:** training dataset is too small or number of training iterations is too small.

**possible fix:** train more iterations on large dataset.

c)

1) c1:

$p_1=3/5, p_2=2/4, r^*=4, c=5, BP=1, BLEU\_c1=1*\exp(0.5*\log(3/5)+0.5*\log(2/4))=0.548$

c2:

$p_1=4/5, p_2=2/4, r^*=4, c=5, BP=1, BLEU\_c2=1*\exp(0.5*\log(4/5)+0.5*\log(2/4))=0.632$

According to BLEU, c2 is better.

I agree with it.

2)c1:

$p_1=4/5, p_2=2/4, r^*=6, c=5, BP=\exp(1-6/5)$

$BLEU\_c1=\exp(1-6/5)*\exp(0.5\log(3/5)+0.5\log(2/4))=0.449$

c2:

$p_1=2/5, p_2=1/4, r^*=6, c=5, BP=\exp(1-6/5)$

$BLEU\_c2=\exp(1-6/5)*\exp(0.5\log(2/5)+0.5\log(1/4))=0.259$

c1 receives higher score. I don't agree with it.

3) In many cases, there are more than one possible translations, even a significant difference between them. If only one kind of translation is used as reference translation, then some good translations will get low scores.

4) advantage:

It is an automatic evaluation method.

Algorithm is simple and intuitive.

disadvantage:

It is depended on the reference translation. A good translation may get low score.

Can't judge the grammar of translation result.