KALLEL Mahdi
mahdi.kallel@telecom-paris.fr

Lab session # 2
ALTEGRAD 2020

12/01/20

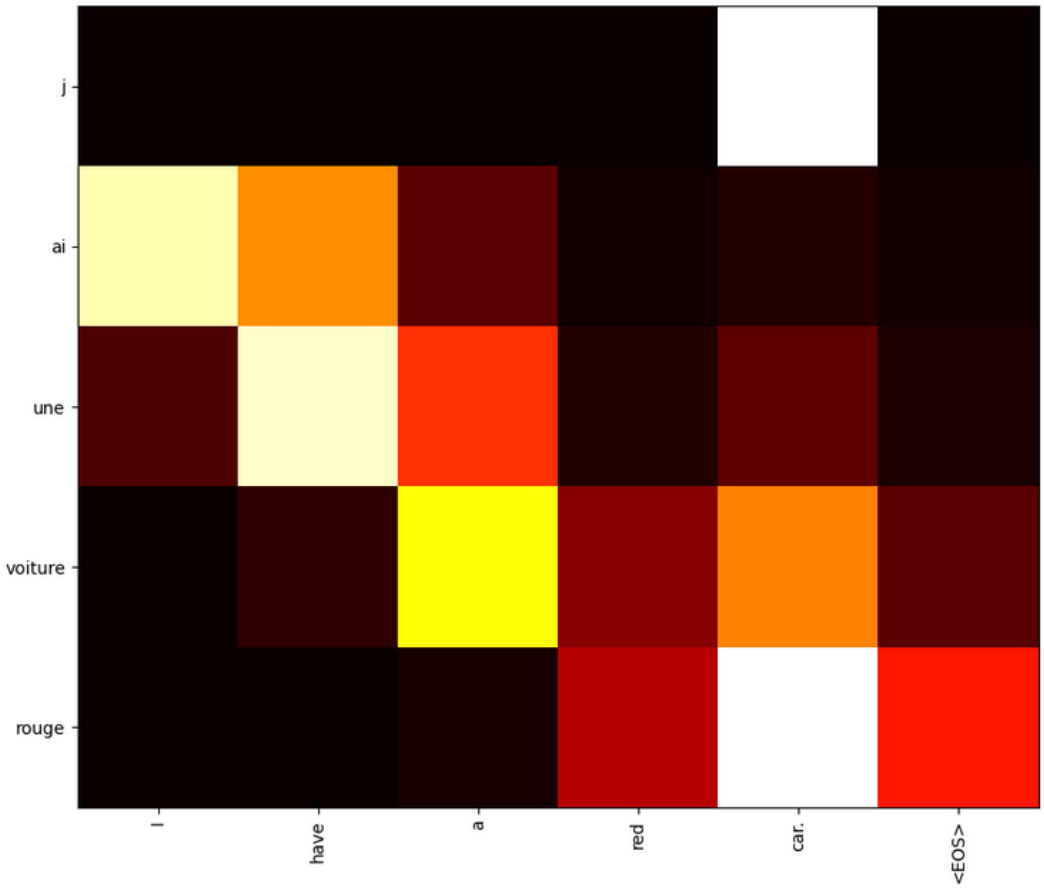
1 Question 1

The greedy decoding strategy consists of picking, at each time step, the token with the highest probability to appear. This introduces bias to our translation as our model is more likely to generate ”easy words” which are the words that are more frequent in our training data set. The model is also negatively biased to not chose rare words.
In our opinion beam search is the most reasonable approach, by keeping track of **B** possible translations at each time step we help the model to privilege overall meaning of the sentence than next token frequency.

2 Question 2

Our model deals well with simple sentences that consisting mainly of **One** independent clause, like ”I have a red car” , but fails and enters into hallucination with compound sentences like ”I can’t help but smoking weed”. In these cases, the translation is not far from right, but the model keeps repeating the same token. The remedy this issue in [3] they propose to use ”coverage vectors” that keep track of how many times a word has been used, this should avoid translating the same word over and over again and force the model to favour less frequent but more meaningful words. In [1] they use ”Input feeding” which consists of feeding the past attentional context to the next time steps. This allows the model to keep track of the which source words have been used during the translation.

3 Question 3



From the heat map above we see that, when translating a word with a certain grammatical function (eg: verb, adjective) the model is paying more attention to it's counterpart in the original sentence. The adjective rouge maps well to red, and the verb have is linked to "ai".
The model is also able to find the the 'gender' of the subject. We see it is balancing it's attention between 'a' and 'car'.
We also notice that the model is able to capture adjective-noun relation as it is focusing on 'red car' when translating to "voiture".

4 Question 4

We report on the pre-trained model translations in the table below.
In the sentence "I did not mean to hurt you. The word "hurt" is repeated over and over. The original translation being "te blesser" it is like the model is realising it had forgotten the subject.
The same goes with "mean" which gets translated to "méchant" many times without according the adjective. Again the model knows it has missed the subject and keeps looping in hope to find it.
In the paper [3] The authors propose the "Bert" architecture. This architecture relays mainly on "Multi-headed self attention mechanisms".
Each head can leverage information from the whole sentence so the context is not limited in the "left" direction and takes into account context from both directions.
Having multiple parallel heads allows for each head to focus on a specific grammatical function.

```
I am a student. -> je suis étudiant . . . . . .
= = = = =
I have a red car. -> j ai une voiture rouge . . . . . .
= = = = =
I love playing video games. -> j adore jouer à jeux jeux jeux vidéo . . . . . .
= = = = =
This river is full of fish. -> cette rivière est pleine de poisson . . . . . .
= = = = =
The fridge is full of food. -> le frigo est plein de nourriture . . . . . .
= = = = =
The cat fell asleep on the mat. -> le chat s est endormi sur le tapis . . . . . .
= = = = =
my brother likes pizza. -> mon frère aime la pizza . . . . . .
= = = = =
I did not mean to hurt you -> je n ai pas voulu intention de blesser blesser blesser blesser blesser blesser . blesser . bl
= = = = =
Help me pick out a tie to go with this suit! -> aidez moi à chercher une cravate pour aller avec ceci ! ! ! ! ! ! ! ! ! ! !
= = = = =
I can't help but smoking weed -> je ne peux pas empêcher de de fumer fumer fumer fumer fumer fumer fumer fumer fumer fumer
= = = = =
The kids were playing hide and seek -> les enfants jouent cache cache cache cache caché caché caché caché caché caché caché
= = = = =
The cat fell asleep in front of the fireplace -> le chat s est en du du pression peigne peigne cheminée portail portail por
= = = = =
She is so mean -> elle est tellement méchant méchant . <EOS>
```

References

[1] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. *The LaTeX Effective approaches to attention-based neural machine translation.*. arXiv preprint arXiv:1508.04025, 2015.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. *Bert: Pre-training of deepbidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.

[3] Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li.
`Modeling coverage for neuralmachine translation` .arXiv preprint arXiv:1601.04811, 2016.5