Théo VIEL
theo.viel.enpc@gmail.com

Lab session # 4
ALTEGRAD 2019

11/24/19

1 Question 1

Our greedy strategy is efficient, but picking the best translation for each symbol works under the hypothesis of independence of the words in a sentence, which is totally false (see Question 4 for a simple example). The best translation of a sentence is not obtained by translating each of its word one by one. Therefore this strategy is too naïve and going for the Beam Search one seems better.

2 Question 2

Our model seems to repeat words during the translation (e.g. : I love playing video games. → j adore jouer à jeux jeux jeux vidéo). As mentionned in section 3.3 of [2], it is because word propositions are made independently. A solution is to introduce a mechanism to keep track of already translated words. The proposed attention mechanism is represented Figure 1.

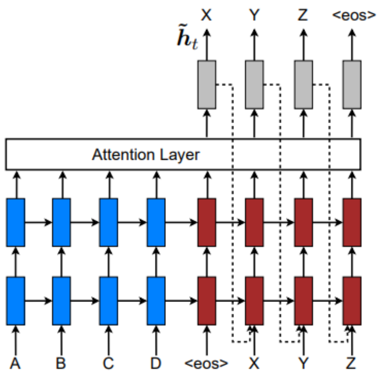


Figure 1: Input-feeding Attention mechanism, see figure 4 of [2]

3 Question 3

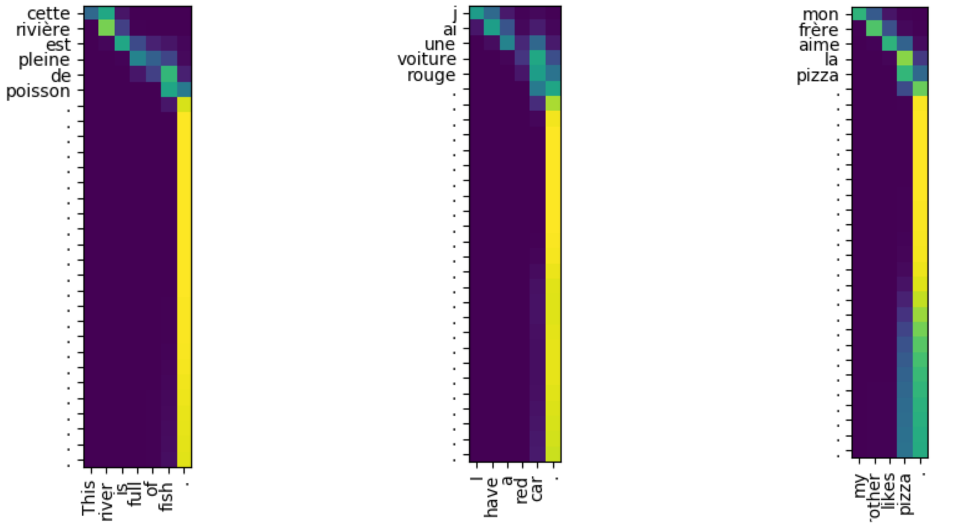
Plots figure 2 emphasize that the attention mechanism manages to capture non trivial grammatical rules:

- Plot 2a demonstrates that ”river” was used for the choice of ”cette” which is necessary to get the gender.
- Plot 2b shows how the attention mechanism reacted to the subject / adjective inversion, and also highlight the previous point.
- Plot 2c shows how ”pizza” was translated to ”la pizza” instead of just ”pizza”.

4 Question 4

- She is so mean → elle est tellement méchant méchant . <EOS>
- I did not mean to hurt you → je n ai pas voulu intention de blesser blesser blesser blesser blesser blesser . blesser . blesser

We can see that although the translations are not perfect, the model managed to capture the different meanings of the word ”mean”. It illustrates that the meaning of a word depends on its context. Indeed, before seeing the word mean, the model has seen words earlier in the sentence, which makes its strength. The alignments figure 3 show that the previous words were taken into account for the translation. BERT [1] pushes the idea of seeing the context of a word by using bi-directional encoders, therefore considering what is before and after the word in the translation.



(a) Accord (b) Subject / Adjective inversion (c) Determinant

Figure 2: Plots of text/targets alignments

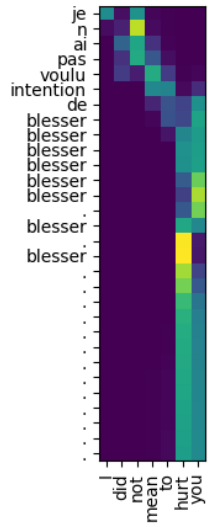


Figure 3: Alignments showing the context understanding of the model.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

[2] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. *CoRR*, abs/1508.04025, 2015.