SIMON Hugo
hugo.simon@telecom-paris.fr

Lab session # 2
ALTEGRAD 2021

23/11/2021

1 Question 1

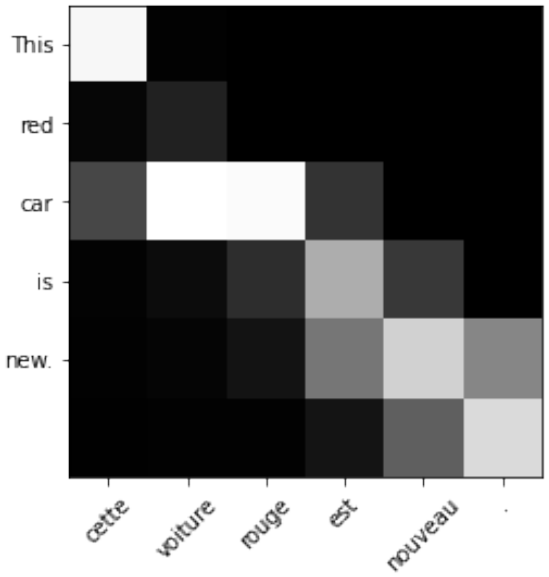
Greedy decoding strategy has the advantage to have a low computational and memory cost. However it doesn't encompass the dependencies among words in sentences by taking the argmax at each step instead of evaluating the probability of a whole sentence or a n-gram and then taking the most probable one. This is what beam search proposes to implement.

2 Question 2

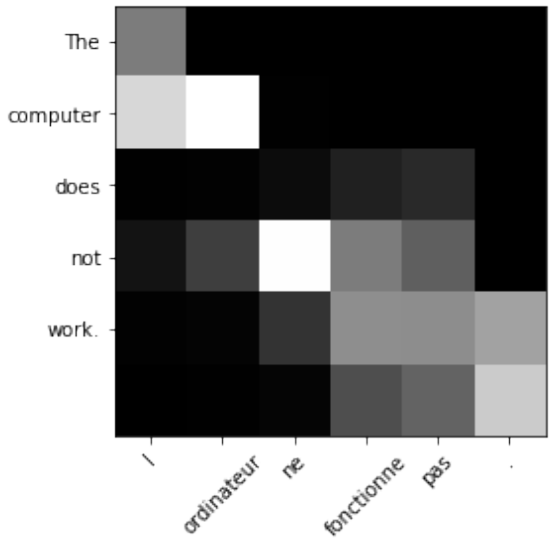
We can see that some words are translated twice and some none. According to Tu et al. [3], this is partly due to the attention-based alignment mechanism of our Neural Machine Translator (NMT). Current alignment ignore previous ones, thus often leads to "over-translation" or "under-translation". This issue can be overcome by using a coverage model such as coverage-based NMT, which recalls at each step which words have been translated or not.

3 Question 3

Fiddling with sentences has shown the difficulty of coming up with cases where adjective and noun inversion is successfully handled by the model. Lack of adjectives in vocabulary and poor translations does not help, and even when the translation is correct, the attention plot does not show clear inversion (see 1a). We also test for negation handling with better results. In 1b, the auxiliary "does" does not rise attention whereas "not" rise attention for both "ne" and "pas.



(a) Attention for adjective-noun inversion



(b) Attention for negation handling

Figure 1: Attention plots for translation examples

4 Question 4

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

Aside from cases of over-translation, our NMT succeeds well in translating ”mean” according to context. Because our NMT only involves unidirectional RNN, one could assume that contextual information about a word in a French or English sentence is contain in previous words. According to Devlin et al. [1] or Peters et al. [2], introducing bidirectional RNN can however significantly improves performances, allowing words to be contextualized by both previous and next words in a sentence.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[2] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations, 2018.

[3] Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. Modeling coverage for neural machine translation, 2016.