Simon Queric
simon.queric@telecom-paris.fr

Lab session # 1
ALTEGRAD 2023

10/15/23

Question 1

The greedy decoding strategy doesn’t require a lot of memory and is fast to compute. The drawback of this strategy is the suboptimality : if the NMT makes a translation mistake at time t, it cannot go back to that mistake and change its estimation of y_t. Once it chose y_t it’s done and it don’t think about the past anymore.

Question 2

There are two major problems of our NMT translations. First, some words are translated multiple times and sometimes a repetition of dots (j adore jouer à jeux jeux jeux vidéo . . .). It is called *over-translation* by [3]. The second problem is that some words are not translated at all (aidez moi à chercher une cravate pour aller avec ceci) : it’s called *under-translation*. These two problems are induced by the *lack of coverage* described by [3]. At each step of decoding, we don’t know if one word has been already translated or not. One solution introduced by [3] is the Coverage Model for NMT. The principle of coverage is to create an array that will keep track of which source words have been translated or not.

Question 3

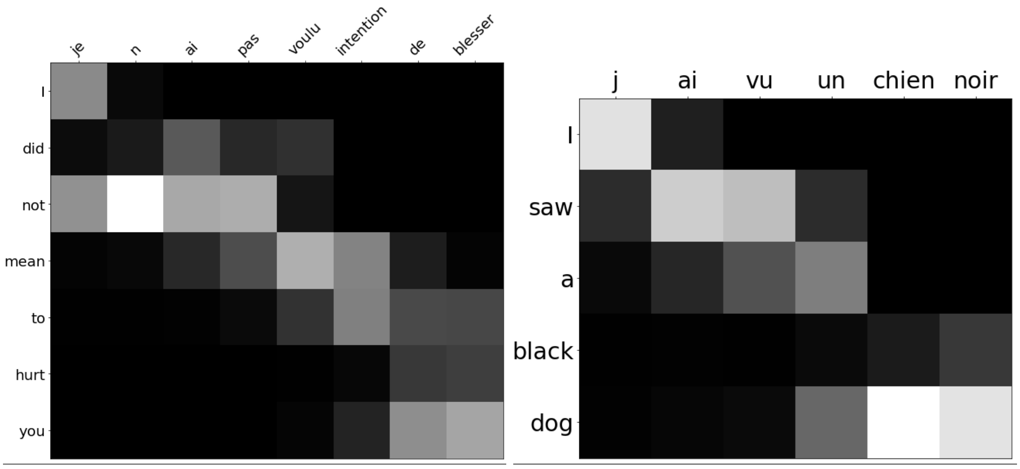

Figure 1: Visualization of source/target alignments of our NTM inspired by [2]. Left : negation is well captured. Right : adjective-noun inversion is well captured.

In the first sentence, the negation is well captured. In english the negation is expressed with the word ”not” while in french there are two words ”n’” (before the verb) and ”pas” (after the verb). It shows a language model is able to infer a grammar rule for the negation in each language. In the second sentence, the adjective-noun inversion is well captured : ”dog” is strongly related to ”chien” and ”black” is related to ”noir”.

Question 4

The translations of the sentences are :

- I did not mean to hurt you. ⟶ Je n’ai pas voulu intention de blesser blesser...
- She is so mean. ⟶ Elle est tellement méchant méchant.

The two translations above illustrate the contextualization property of language models. The word ”mean” can be translated either by the verb ”vouloir” (first sentence) or by the adjective ”méchant” (second sentence) in french (but also by the verb ”signifier” and by the noun ”moyen” depending on the context). It shows the NMT takes into account the linguistic context in which the word is used. This property is called *model polysemy*. The model BERT proposed by [1] is based on the fact that the meaning of a word is determined by the linguistic context. In a language model, each token embedding depend on the context of the source 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] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation, 2015.

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