NAOUMI Salmane - KILANI Al Houceine
salmane.naoumi@gmail.com
houcinekila@gmail.com

1 Questions :

1.1 Question 1:

This decoding technique seems to be very efficient computationaly as underlined in the slides. But the problem is that sometimes, when following the conditional dependency path, it produces very long sentences before the eos token will be returned by the argmax. Thus it is in this case "suboptimal" is the sense that it yields big sentences where as there a lot of other smaller sentences that are better and make more sense.

1.2 Question 2:

```
= = = = =
 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 de jeux jeux vidéo . . . . . . . . . . . . . . . . . . . . . . .
= = = = =
 This river is full of fish. -> cette rivière est plein 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 dire dire faire faire blesser de blesser . . blesser . . . . . . . . . . . . . . . .
= = = = =
 She is so mean -> elle est tellement méchant de dire dire dire . <EOS>
= = = = =
 Help me pick out a tie to go with this suit! -> aidez moi à chercher une cravate pour aller avec ceci ! ! ! ! ! ! ! ! ! ! ! ! ! ! ! <EOS>
= = = = =
 I can't help but smoking weed -> je ne peux pas empêcher de de fumer fumer fumer fumer fumer fumer fumer . fumer urgence volonté volonté volonté
volonté volonté volonté . . . . . . .
= = = = =
 The kids were playing hide and seek -> les enfants jouent cache cache cache cache caché et caché caché caché caché caché caché caché caché caché
caché caché caché dentifrice perdre risques rapide caché risques rapide caché risques
= = = = =
 The cat fell asleep in front of the fireplace -> le chat s est endormi au événement de événement portail de oiseaux . embouteillage oiseaux
oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux oiseaux
```

Figure 1: Translations given by the pretrained models finetuned for 4 more additional epochs

The problem in our translations is that there seems to be some repetition. Indeed this attention model fails to "take advantage of past alignment information". That means that when we translate a word at a step t, we do not know if we have translated that word before. To solve this problem, many technique have been proposed such as "input-feeding"(where attentional vectors are concatenated with input of the next state) and "coverage"(where a word previously translated has a lower probability to be translated further).

Another problem that is happening is that, since we use a prediction in time $t-1$ to predict the word at time t, once we make a mistake in the prediction, the hidden states of the model will be updated by a sequence of wrong predictions, errors will accumulate, and it is difficult for the model to learn from that. One way to solve this is to use what is called "teacher forcing" where we ommit previous predictions and force ground truth labels to the next input layer of the RNN. This works well during training, but during inference, since there is usually no ground truth available, the RNN model will need to feed its own previous prediction back to itself for the next prediction. Therefore there is a discrepancy between training and inference, and this might lead to poor model performance and instability.

1.3 Question 3:

Our Seq2Seq model enables to retrieve attention scores between source and target inputs. Here we illustrate some of the resulting examples we got after retraining the pre-trained model.

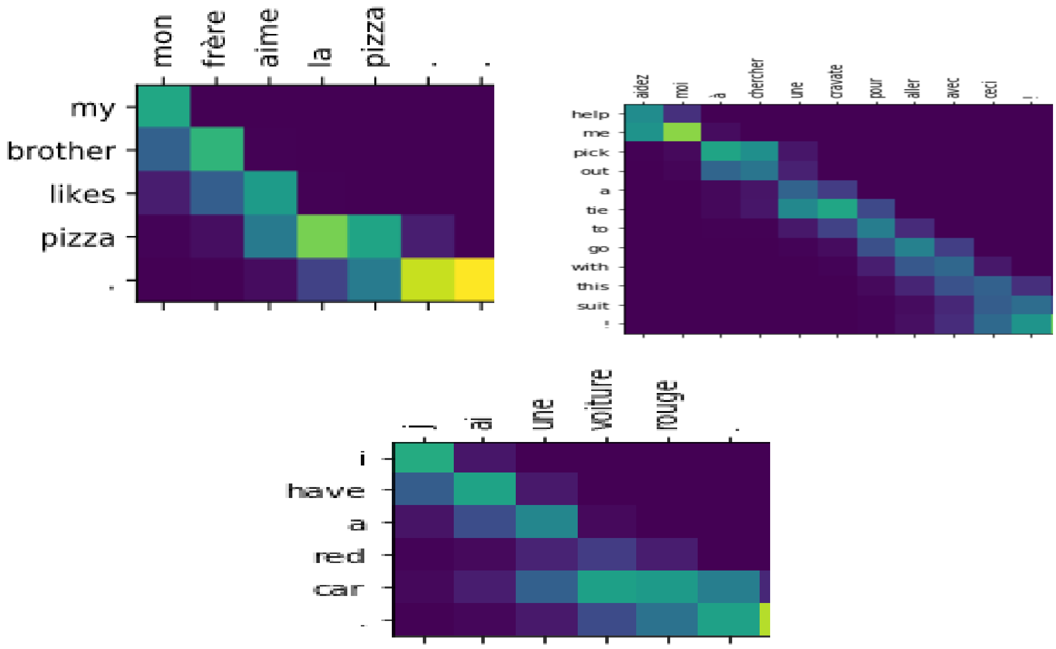

Figure 2: Examples of source/target alignments using trained model

In the first example, source and target alignments is showing that **pizza** is translated to **la pizza** and that scores are high for adjacent words. From the third example, the model has captured the need to inverse **voiture** and **rouge** which is quite impressive and show where these scores seem to be useful useful.

1.4 Question 4:

The two sentences were translated as follows:
I did not mean to hurt you $\Longleftrightarrow$ je n ai pas dire dire faire faire blesser de blesser.
She is so mean $\Longleftrightarrow$ elle est tellement méchant de dire dire dire.

Even if the translation is not totally accurate, the model managed to translate the word **mean** in the context of each sentence showing how powelful this way of representation based on context vector is way more useful than simple word2vec or GloVe.

References