CHAUVIN Paul
paulchauvin97@gmail.com

Lab session # 1
ALTEGRAD 2022

11/14/22

1 Question 1

1. ”A structured Self-Attentive Embedding” proposes a way to improve the self-attention mechanism by introducing a bi-directional LSTM before feeding the attention layer. So as before we get for every word $w_{i,t}$ in sentence i embedding $x_{i,t}$ that we feed as input in a bi-directional LSTM :

$$\overrightarrow{h_{it}} = LSTM(x_{i,t}, \overrightarrow{H_{i,t-1}})$$
$$\overleftarrow{h_{it}} = LSTM(x_{i,t}, \overleftarrow{H_{i,t+1}})$$

The concatenated output annotation $h_{it} = [\overrightarrow{h_{it}}, \overleftarrow{h_{it}}]$ summarises the information of the whole sentence centred around the word wit. In the attention-scoring function we could interpret the annotation as following :

- $\overrightarrow{h_{it}}$ is the key
- $\overleftarrow{h_{it}}$ is the query
- h_{it} is the value

We now cross correlate information in the sentence from both side of the temporal axis.

2 Question 2

In the paper ”Attention is all you need”, authors proved that Transformers architecture achieve better results on two machine translation tasks than the state of the art (of 2016, which included RNNs or CNNs in part of the architecture). Furthermore, they are highly parallelizable thanks to its multi-head architecture (linear projection of annotations which can be trained in parallel) and train much faster since model don’t need to back-propagate through time-steps, avoiding the risk for vanishing or exploding gradients.

3 Question 3

Here are the coefficient obtained in ”my_review” with the corresponding values :

7.89 There ’s a sign on The Lost Highway that says : OOV SPOILERS OOV (but you already knew that , did n’t you ?)
5.64 Since there ’s a great deal of people that apparently did not get the point of this movie , I ’d like to contribute my interpretation of why the plot
4.48 As others have pointed out , one single viewing of this movie is not sufficient .
5.65 If you have the DVD of MD , you can OOV ’ by looking at David Lynch ’s ’Top 10 OOV to OOV MD ’ (but only upon second
3.53 ;) First of all , Mulholland Drive is downright brilliant .
3.67 A masterpiece .
5.56 This is the kind of movie that refuse to leave your head .

The results here don’t appear very consistent regarding the type of message and the type of score that we are getting.

4 Question 4

A limitation to the HAN architecture is that each sentence is encoded independently and in isolation. Hence, we could lack some information regarding the context of the sentence. When a sentence gets a representation it is done while ignoring all the other neighbouring sentences. Hence most of the attention is directed towards salient cues as opposed to volutional cues which should be infered by a better designed attentional scoring mechanism.