Pradel Guillaume
gpradel88@gmail.com

Lab session # 1
ALTEGRAD 2019

10/14/24

1 Question 1

The attentional vector s focuses on a specific aspect of the sentence, such as a group of related words. Thus, s is focusing on the semantic of the sentence. Nevertheless, s can't focus on several different aspects of the sentence to find the overall meaning and it can miss understand some sentences.
For instance : "It's raining cats and dogs !". In this sentence, the algorithm will surely focus on 'cats' and 'dogs' and will associate the meaning of the sentence with an animal context. But here, we have to take the sentence as a whole to understand that the sentence means 'It rains a lot'. The article [1] suggest to compute various vector s in order too find extract different information. Then, to compute multiple hops of attention.

2 Question 2

The main problem of reccurent computation is the *sequential nature* of this model. In fact, the reccurent computation generate a sequence h_t based on the previous sequence h_{t-1}. Thus, as it is explain in [3] that parallelized computing is not possible and long sequences of words have a huge impact on the memory.

3 Question 3

Let's see how the model will work with an excerpt from the movie Fight Club, 'Tyler Durden Speech'.

```
2.12 We are the middle children of history, w
8.81 We have no great war, or great depressio
16.21 The great war is a spiritual war
21.93 The great depression is our lives
4.96 We were raised by television to believe
3.44 And we're learning that fact
```

Figure 1: Sentence attention coeff

We will focus on sentence attention coefficients. We can see a few of them above, all the other are very low (around $5.$). But here, we can witness that 2 sentences have a way higher coeff. The sentences 'The great war is a spiritual war' and 'The great depression is our lives'. First, let's see the original vectors that represent the sentences :

$$[22, 100, 525, 9, 6, 3887, 525, 0, ..0]$$

$$[22, 100, 5694, 9, 293, 471, 0, ...0]$$

These sequences have something in comon that can be interesting to explore, both contains the word 'great'. And, as the model has been trained on movie reviews, it is possible that the word 'great' has a great weight in these sentences. We can try a little experiment, to test that. Let's try to had a sentence : 'Sport is a great outlet for excess energy'. By adding this sentence, the sentence attention of both original sentences decreased respectively to 13.01 and 18.45. And the attention of the sentence : 'Sport is a great outlet for excess energy' went up to 15.05.
Let's go in depth and check the word cofficients we obtain for 'The great war is a spiritual war' and 'The great depression is our lives'.

```
= = = =
('great', 0.11292720521932305)
('great', 0.08650412836090764)
('depression', 0.05398333802350083)
('or', 0.05161758225756222)
('We', 0.03736113338537351)
('war,', 0.03432394627907602)
('no', 0.034036036725722696)
('have', 0.031192281483220494)
= = = =
('spiritual', 0.1724340723844905)
('great', 0.11493964732004916)
('war', 0.09650779604823169)
('war', 0.04263910347543466)
('a', 0.04018178350274545)
('is', 0.03583552607226085)
('The', 0.03340672542390204)
= = = =
```

Figure 2: Sentence attention coeff

As expected, the word 'great' has a huge weight in this model, and it is at the root of the weight of these sentences.

4 Question 4

HAN architecture encode each sentence independently of the others. Which lead to a loss of overall context in the document. As presented in [2], for instance, when the same sentence is repeated.

References

[1] Zhouhan Lin, Minwei Feng, Cícero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. *CoRR*, abs/1703.03130, 2017.

[2] Jean-Baptiste Remy, Antoine J.-P. Tixier, and Michalis Vazirgiannis. Bidirectional context-aware hierarchical attention network for document understanding. *ArXiv*, abs/1908.06006, 2019.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.