SIMON Hugo
hugo.simon@telecom-paris.fr

1 Question 1

A basic attention mechanism is prone to generating weights not varying significantly enough to have efficient weighting across the sentences of the text. Hence, we can improve the model by adding a penalisation term. Lin et al. [1] notes that in this particular case, it is much more appropriate to use a term based on the Frobenius matrix norm rather than on the Kullback-Leibler divergence.

2 Question 2

According to Vaswani et al. [3] on Transformer, the main motivations for replacing recurrent operations with self-attention are: the considerably lower computational cost for each layer, along with better paralleling capabilities, and the lowering of the distance between long-range dependencies.

3 Question 3

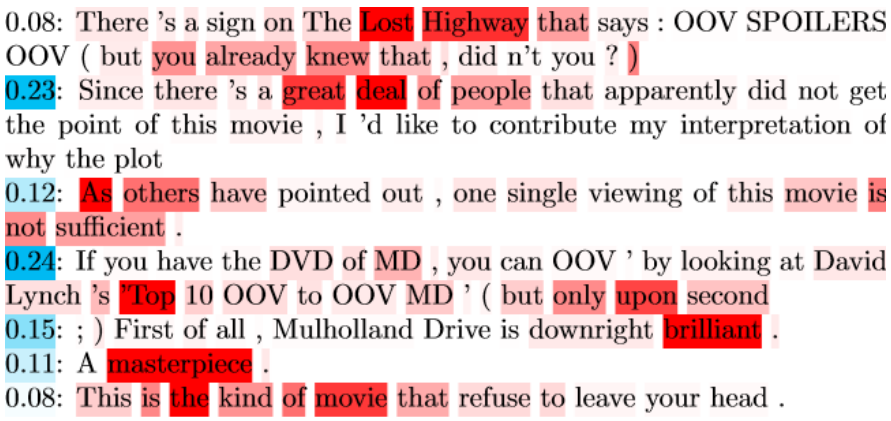

Figure 1: Sentences and word attention from HAN

One can see in 1 the encoding performed is not ideal. High word attention is set on what seem irrelevant words such as "Lost", "Highway", "that". Moreover, high sentence attention is set on what seem irrelevant sentences such as the second one, and low sentence attention is set on what seem relevant sentences such as the last 2.

4 Question 4

Limitation of Hierarchical Attention Network (HAN) is that attention-based encoding of sentences is independent between sentences. This means that a HAN can focus its attention on the same in-sentence structure without noticing relation of this structure across sentences. A Context-Aware Hierarchical Attention Network (CAHAN) proposed in Remy et al. [2] allows to adapt attention to the context of sentences, preventing high attention on redundant information.

References

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

[2] Jean-Baptiste Remy, Antoine Jean-Pierre Tixier, and Michalis Vazirgiannis. Bidirectional context-aware hierarchical attention network for document understanding, 2019.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.