Homework 4: Machine Translation

CS 1470/2470

Due November 5, 2019 at 11:59pm

1   Conceptual Questions

1. What is the purpose of the positional encoding in the Transformer architecture? (2-3 sentences)

2. What are the limitations of RNNS that Transformers solve? (3-6 sentences)

3. Consider the parameters for two different attention heads. Is it necessary that they be initialized randomly, or could we just start them all with the same vector of values? (2-4 sentences)

4. In an encoder, that has attention from right to left, and left to right) is the attention word 5 pays to word 10 the same as that word 10 pays to word 5? (3-5 sentences)

5. (Optional) Have feedback for this assignment? Found something confusing? We'd love to hear from you!

2   Ethical Implications

In lecture, we discussed the environmental impacts of deep learning, and saw how transformers are some of the most energy-intensive models to train. We also saw a few suggestions for how the field could keep those factors in mind moving forward.

1. Network pruning is an active area of research within the field of deep learning. Why does the strategy described in lecture "work"? How does it have the potential to reduce the long-term environmental effects of deep learning? (2-3 sentences)

2. One suggestion for dealing with this is that researchers should report training time and hyperparameter sensitivity and consider computational cost as a factor in evaluating or pursuing new work.

    (a) What, if any, are the disadvantages of the research community moving in this direction?

(b) What is your evaluation of how likely or feasible this suggestion is? Do you think individuals are likely to start considering these considerations on their own? Why/ why not?

3. How, if at all, does learning/thinking about the environmental consequences of the field affect how you approach deep learning? If so, how? If not, why not? (2-4 sentences) Again, no right or wrong answers here – we are looking for evidence of critical thinking, and the application of your personal beliefs and values to this problem.

3 CS2470-only Questions

1. What requires more parameters: single or multi-headed attention? Explain. Does this mean one necessarily trains faster than another? (3-5 sentences)

2. For the last homework, we asked you to consider convolutional architectures for language modeling, and to weigh their trade-offs against RNN-based architectures.

   Transformers can also be used for language modeling. In fact, they are current state of the art.
   (see https://openai.com/blog/better-language-models/).

   How are transformer-based language models similiar to convolution? What makes them more suited for language modeling? (5-10 sentences)