CS 4644/7643: Deep Learning
Fall 2022
HW3 Solutions

Arthur Scaquetti do Nascimento

March 12, 2024

4.7)

- Key contributions: This work proposes a technique for pretraining sequenc-to-sequence models with denoising objectives for learning general-purpose representations, which was thought for NLP, but is a cornerstone to transfer to other modalities. It explores different techniques for denoising, is by design able to handle different NL processing tasks, and was the benchmark to be beaten when it came out.
- Strengths:
  - It can handle different tasks without task-specific modifications, which leads one to infer that its generalization capabilities are very solid.
  - The method proposed is shown to be very robust by nature, i.e., the denoising foundations of it make it so that it is able to handle corrupted data,
- Weaknesses: As expected, BART is extremely computational complex. On top of that, I didn't find an effort into pinpointing the bottleneck, what aspect of BART contributes more to its computation complexity. Also, we still don't find interpretability or explainability investigated.

4.8) Personal takeaways:
While reading, it bothered my a lot that the latest research in this large models filed oversees the causes of computation complexity. Yes, there is an obvious direct correlation between the number of parameters of a model and its complexity, but I am starting to wonder if there is something more to it than that.
As in what methods, operations or techniques take more compute? If it is the denoising process, which part? I have enver seen an ablation study like this myself. Moreover, it get me thinking about the first PS: up to what point it is the architecture's fault? I have no idea how to answer that, but it's something that got me thinking.