11-877 Advanced Topics in Multimodal Machine Learning

Week 4: Pretraining paradigm

Due date: 11PM EST, Wednesday, Feb 9 2022

Submission: https://forms.gle/zYRtYTzb5Snu5uJo9

We designed the reading assignments to help you prepare for the live discussions. Discussion probes were drafted related to this week's topic. These were written to help conceptualize the problem and guide your thought process. Take the time to read them first. The goal is not to answer each of these questions and probes individually, but they are meant to be taken as a whole. We also selected research papers relevant to this topic. Required papers should be read completely. Suggested papers should at least be skimmed. The purpose of the reading assignment is to start your critical thinking process, so your responses should demonstrate constructive thoughts, with a good understanding of the current research in this area, and expressing your own insights.

Your response to this reading assignment should be submitted in the online Google Form (see link above). Your response should consist of four main components:

(1) **Scouting:** As you start thinking about the discussion probes, it is always good to also scout papers, blog posts and other resources related to the topic. We ask that you search for related resources and share with us 2 extra links to these new resources. For each extra link, include 1-2 sentences explaining the value and relevance of this extra resource.

(2) **Reading notes:** As you read the required papers, suggested papers and the extra resources you scouted, please write down at least 4-6 notes related to the discussion probes. Each note should be 1-3 sentences long. These can be empirical results you observed, ideas or theories expressed by other researchers, or any interesting fact that is worth noting when summarizing your reading.

(3) **Your thoughts:** Separate from your reading notes, we ask that you reflect more holistically about the discussion probes. Please write 3 discussion points you would like to share on this topic. Each discussion point should be one paragraph (3-5 sentences). These discussion points should go beyond the reading papers, and try to address as many aspects of the discussion probes as you can. We do not expect that you answer all discussion probes. For example, it would be ok to focus on only 1 or 2 probes if these bring the most ideas and thoughts for you.

(4) **Clarification requests [OPTIONAL]:** Please take a moment to suggest parts of the papers where clarifications would be useful. Try to be as specific as possible in your clarification requests. These requests will be shared with the Reading Leads in charge of creating a short presentation for the beginning of Friday course and answering other requests directly on Piazza.

Week 4 discussion probes:

- Is large-scale pre-training the way forward for building general AI models? What information potentially cannot be captured by pre-training? What are the risks of pre-training?
- What are the types of cross-modal interactions that are likely to be modeled by current pre-training models? What are the cross-modal interactions that will be harder to model with these large-scale pre-training methods?
- How can we best integrate multimodality into pre-trained language models? What kind of additional data and modeling/optimization decisions do we need?
- What are the different design decisions when integrating multimodal information in pre-training models and objectives? What are the main advantages and drawbacks of these design choices? Consider not just prediction performance, but tradeoffs in time/space complexity, interpretability, and so on.
- How can we evaluate the type of multimodal information learned in pre-trained models? One approach is to look at downstream tasks, but what are other ways to uncover the knowledge stored in pre-trained models?

Required papers (you should read these papers in detail)

- https://arxiv.org/abs/2102.00529
- https://arxiv.org/abs/2106.13884

Suggested papers (you should skim through these papers, at the minimum)

- https://arxiv.org/abs/2102.02779
- https://arxiv.org/abs/2112.04482
- https://arxiv.org/abs/2103.05247

Other relevant papers:

- Survey: https://arxiv.org/abs/2108.07258. Particularly sections 2.2 vision, 2.3 robotics, 2.5 interaction, 3.1 healthcare and biomedicine, 4.1 modeling.
- https://arxiv.org/abs/2109.10246
- https://arxiv.org/abs/2005.07310
- https://arxiv.org/abs/1908.05787
- https://arxiv.org/abs/2102.12092 and https://openai.com/blog/dall-e/