11-877 Advanced Topics in Multimodal Machine Learning

Week 2: Cross-modal interactions

Due date: 11PM EST, Wednesday, Jan 26 2022

Submission: https://forms.gle/zquUqvyaED1BFt9NA

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 three 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 1 or 2 extra links to these new resources.
(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, in light of your readings. Please write 3 discussion points you would like to share on this topic. Each discussion point should be one paragraph (3-5 sentences). You should try to address as many aspects of the discussion probes as you can, but your goal is not to answer all of them. If one discussion probe brings more ideas and thoughts for you, you are welcome to have 2 or 3 points on this probe.

Week 2 discussion probes:

- What are the different ways in which modalities can interact with each other in multimodal tasks? Can we formalize a taxonomy of such cross-modal interactions, which will enable us to compare and contrast them more precisely?
- What are the design decisions (aka inductive biases) that can be used when modeling these cross-modal interactions in machine learning models?

- What are the advantages and drawbacks of designing models to capture each type of cross-modal interaction? Consider not just prediction performance, but tradeoffs in time/space complexity, interpretability, and so on.
- Given an arbitrary dataset and prediction task, how can we systematically decide what type of cross-modal interactions exist, and how can that inform our modeling decisions?
- Given trained multimodal models, how can we understand or visualize the nature of cross-modal interactions?

Required papers (you should read in details these papers)

- Additive interactions: https://aclanthology.org/2020.emnlp-main.62/
- Grounding interactions: https://aclanthology.org/2020.acl-main.469.pdf (short paper)

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

- Multiplicative interactions: https://openreview.net/pdf?id=rylnK6VtDH
- Cooperative interactions: https://arxiv.org/pdf/2112.12337.pdf
    - Feel free to skip the equations and derivations. It is suggested to read the introduction and motivation for their algorithm (section 1). Other sections of interest are Sections 2.4, 4 and 7.2.
- Visualizations and ablations: https://arxiv.org/abs/2109.04448

Other relevant papers:

- Visualizations and ablations: https://arxiv.org/abs/2012.12352