11-877 Advanced Topics in Multimodal Machine Learning

Week 15: Generalization, Low-resource & Robustness

Due date: 11PM EST, Wednesday, April 27 2022

Submission: https://forms.gle/Xq4TCY4METSDpQ128

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 express 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's course and answering other requests directly on Piazza.

Week 15 discussion probes:

- One general claim is that pre-trained models can help with low-resource settings (e.g., few-shot fine-tuning). What are the multimodal problems where the paradigm of pre-training and fine-tuning may not generalize? What are the technical challenges?
- What are new research paradigms that should be explored to address the challenges of multimodal low-resource problems? Can you propose a taxonomy of the challenges that should be addressed to make progress in this direction, for low-resource modalities?
- How can we develop new models that generalize across many modalities, going beyond only 2 or 3 modalities? What are the tradeoffs between modality-specific multimodal models and general-purpose multimodal models?
- What are the commonalities and underlying principles shared across diverse modalities and tasks that can enable good generalization? In other words, what are the pre-requirement for generalization to succeed?
- What are the limits of generalization? In other words, in which cases is generalization across modalities and tasks not possible due to possibly to data heterogeneity or some other reasons? What are these scenarios where generalization may not be possible?
- How can we potentially perform generalization of multimodal models in the absence of explicit alignment (e.g., paired data) between modalities? How can we tackle the challenges of learning cross-modal interactions, alignment, reasoning, etc?
- One other aspect of generalization is with real-world settings where noise is present and modalities may be even missing. How can we robustly handle these noisy situations? How can multimodal help? Can multimodal also make these noisy situations harder?

Required papers (you should read these papers in detail)

- https://arxiv.org/abs/2204.00598 (you can focus on the first 3 sections)
- https://arxiv.org/abs/2201.04309 (the proof in section 4.2 could be skipped)

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

- https://arxiv.org/abs/2010.12831
- https://arxiv.org/abs/1811.10787
- https://arxiv.org/abs/2111.07991
- https://openaccess.thecvf.com/content_cvpr_2017/html/Tran_Missing_Modalities_Imputation_CVPR_2017_paper.html
- https://dl.acm.org/doi/pdf/10.1145/3394486.3403234

Other relevant papers:

- https://arxiv.org/abs/1811.08615
- https://arxiv.org/abs/2011.08899
- https://arxiv.org/abs/2107.07502