Language Agents in Interaction

CS 294-258, Spring 2024
Tuesday/Thursday 3:30-5pm

Intro Questionnaire

- Find the link in Edstem!
- Please fill out before the second lecture
- We might try to move the class to BWW if there is interest

Class Format

- **Meetings**
  - Pretty much every meeting will be discussion of a paper assigned at least a week in advance
  - We may have some guest lecturers throughout the semester, but this is TBD
  - Last couple meetings reserved for project presentations
- **Project**
  - Group project
  - Deliverables: project proposal (partway through semester), final report (~paper), final presentation (~conference talk)
  - More details on the project will be posted soon

Readings

- Every week will have a topic
  - This week: “interaction” (broad intro topic)
  - Next week: word learning
  - Weeks after: you decide!

- **Tuesday meeting:** discussing an foundational, and/or interdisciplinary work (usually an older paper)

- **Thursday meeting:** discussing a cutting-edge, usually computational paper

Proposing Topics

- In Edstem there's a pinned post where you can propose topic suggestions and vote on other suggestions: https://edstem.org/us/courses/52873/discussion/4122911
- Please add your own topics as well!!!
- I'll choose new topics from this list!
- Vote by hearting top-level comments

Role-Playing Paper Reading

- Jacobson and Raffel proposed this kind of reading group a couple years ago (https://colinraffel.com/blog/role-playing-seminar.html)

- **Main point:**
    - For every paper, you get some role, and you should read the paper with that role in mind
    - During in-class discussion, we will discuss papers for each role in order
    - You will have the same role each week (one exception), and it will be mostly randomly assigned

- **Role representative:**
    - You'll be assigned as a role representative about once every other week (depending on enrollment)
    - When you are the role representative, you present the role's findings for a particular paper to the rest of the class
    - I (Alane) will always be one of the role representatives

Before Meetings

- Check your role for the week
- Check if you will be a representative (and if you can’t make it, let me know ASAP)
- Read the papers with your role in mind
- Start prepping material for the presentation
- If you are a role representative, be proactive about reaching out to other members of the role

During Meetings

- **First 10 minutes:** in-role discussions; role representative will lead a discussion with other people who have the role, and prepare the slides for presentation

- **Most of class:** role representatives will present the role to the rest of the class and lead a discussion; ~10 min each

- **Last 10 minutes:** breakout groups split by pre-assigned group numbers (per day) that mix roles together, for wrap-up discussion of the paper

After Meetings

- Post a really short summary of the discussion/paper on the Edstem thread for the paper

- Format: ~tweet-length comment/question on something that was inspiring/intriguing for you

- Highly recommend posing this as an open-ended research question, or bringing up a particularly interesting example

Role Assignments

Name	Role	Group (January 23)	Role Rep? (January 23)	Time Traveler? (January 25)	Group (January 25)	Role Rep? (January 25)
Alane Suhr	Original Author	1	Yes	Yes	2	Yes

Roles
(e.g., LSTM + Transformers papers)

- **Original author**
- **Time traveler** → why didn't LSTM designers end up with a Transformer instead?
- **Archaeologist** → what were original applications of the LSTM? What was the inspiration for Transformers?
- **Scientific peer reviewer**
- **Sociologist** → what kind of compute do Transformers require?
- **Industry practitioner** → what non-language applications might Transformers have?
- **Academic researcher** → how can you improve on the Transformers architecture?

Grading

- Subject to change, numerical grade to letter grade will be determined near end of semester
- 50% course participation
  - 25% presentations as role representative
  - 25% after-meeting summaries
- 50% course project
  - 15% proposal
  - 20% final project report
  - 15% final presentation

Misc Policies

- **Attendance:** you will need to attend
  - If you have a conflict let me know, it's likely you will have to drop
  - Attendance gauged by post-discussion summaries
    - Totally understandable to miss a day or two, due to illness, traveling, etc. Just let me know (especially if you are assigned as a role representative).

- **Communication:** use Edstem but you can also email me, just put 294-258 in the subject

- **COVID-19:** if you feel sick, do not come to class. Email me when you are feeling better. I strongly encourage students to wear masks.

Language Agents in Interaction

- **Agent:** something that changes the state of the world by taking actions

- **Language:** a special subset of one's action space that recruits other agents towards some goal

- **Language agents:** an agent that uses, or is a recipient of, language (the authoritative language agent: a human!)

- **Interaction:** a multi-turn dynamic process where two or more agents respond to each other's actions

Concept Representation

- We divide the world into mostly-discrete but still flexible abstractions
- How are the abstractions we maintain influenced by our general behavior in the world? By the language we use?
- How are these abstractions recruited in regular language use across different modalities (e.g. vision vs. hearing)?
- How do abstractions differ across members of a language community?
- How are abstractions related to one another (e.g., as a hierarchical taxonomy)?
- How is this different in language models, particularly multimodal language models?

Knowledge

- Beyond labels for objects, attributes, and actions, we know facts, processes, skills
- We know how to use our knowledge of the world
- We maintain uncertainty over this knowledge, and learn through interaction with the world
- How do humans represent knowledge, including uncertainties, strongly-held beliefs, and consistencies thereof?
- How does our representation of knowledge engage with our language use in conversation?
- How are our uncertainties and beliefs surfaced through language use?
- How do language technologies represent knowledge?

Production and Comprehension

- Two sides to language use: comprehending an utterance someone made, and producing our own utterances
- How do we resolve a speaker's intent from their utterances?
- How do we turn intent into an utterance directed towards a listener?
- How do these processes interact with one another?
- Are these processes anything like what an LLM is doing?

Pragmatics and Theory of Mind

- When we use language, we don't just take into account the literal meaning of utterances we encounter (or generate)
- We also take into account how the interlocutor is representing the world and their language use
- How do we model interlocutors (and our uncertainties of those models)?
- How do we use that model to most efficiently, yet effectively, recruit our interlocutors?
- How can we use language to move our representations of the world closer together?
- Can / do modern language technologies model interlocutors?
- How would we evaluate such a thing?

Ambiguity and Uncertainty

- Language is intrinsically ambiguous
- A huge part of interactive language use is resolving uncertainties held about one another and the world
- How do we maintain ambiguity in interactions?
- What kinds of uncertainties are we comfortable with, and which must we resolve?
- How do we functionally use ambiguity to get things done with language?
- What is the role of ambiguity in interaction and language change?
- Can language technologies maintain ambiguity in conversation with human interlocutors?

Reference

- To successfully interact with another language agent, we must share some common ground
- In a situated interaction, this often requires making and resolving references to things outside the language space, e.g., the surrounding environment
- How do we generate and resolve references grounded in situated contexts?
- How do we make references to things that *are not* present in the current situated context, e.g. hypothetical events?
- How do we refer to abstract concepts?
- Has this been solved by LLMs/VLMs?

Spatial Language

- References to space and spatial relations are proposed to form the basis of metaphor (Lakoff and Johnson, 1980, Metaphors We Live By)
- Spatial relations are highly contextual and nuanced
  - The language we use to refer to spatial configurations and motion depends on perception, estimates of the interlocutor's perception, and conventions we've formed in interaction
  - How do people integrate perception of space and motion with language use in interactions?
  - What's missing from current vision-language models to handle spatial language as precisely as people do?

Real-Time Language Use

- When we hold a real-time situated conversation, we are coordinating very precisely with our interlocutor
- We coordinate on turn-taking, backchannel feedback, prosody, etc.

- What underlying structures govern our participation in these conversations?
- What other paralinguistic features contribute to meaning?
- How far are we from building language technologies that precisely, yet flexibly, participate in these kind of interactions with humans?

Discourse and Dialogue

- In long conversations, or over multiple interactions with the same person, we form conventions and common ground
- We are able to intentionally add to common ground / propose new conventions, and store these conventions in long-term memory

- How do our utterances engage with broader conversational context?
- What are the underlying structures of conversation, e.g. as we shift from topic to topic?
- How can we maintain a consistent representations of conversations over a long period of time?
- Where do current language technologies fail in this problem?

Word Learning (next week!)

- Children learn new words for objects, attributes, and actions super quickly
- We fine-tune the meaning of words through interaction with other language users

- How do children learn words efficiently without “direct supervision”?
- What principles and initiatives do children take to learn words?
- How do we bootstrap refinement of perception/action and word learning?
- How do the words we learn influence our representation of the world?
- Can we use these principles of word learning to train models with less data, particularly in multimodal contexts?

Language Acquisition

- Language is more complex than learning labels for objects, attributes, and actions
- We also must acquire complex linguistic structure that forms the basis of communication, e.g., syntax
- However, children don't have access to nearly the same amount of data as a LLM does!
- How do children learn language with miniscule amounts of language data?
- What role does interaction play in this process?
- How can we take inspiration from theories of child language acquisition to design more efficient learning algorithms?

Artificial Language Agents

- Never before have we encountered systems that appear to use language like any other language agent (human)
- How do individuals think about these agents during interaction?
- How are we changing the way we use language?
- What are our expectations of these systems?
- Is the existential crisis of automated language users really a new crisis?

Language and Society

- Language is a political tool
- The linguistic structures we use signal our membership of (or independence from) social groups
- The metaphors we propose influence our perception of the world
- We argue about what words mean to influence how they are applied in practice
- How does language engage with a broader social context in light of broader social conflicts or issues?
- How do language technologies represent social norms and values?

Language Variation

- Language is incredibly diverse
- Diversity in local contexts: e.g., talking to friends vs. family
- Diversity in global contexts: i.e., linguistic diversity across a wide range of dimensions
- Along what dimensions can we characterize languages (and thus their diversity)?
- What structures are fundamentally shared across languages? Which are not?
- How do languages differ from one another, and why?
- How does multilingualism influence an individual language user's relationship with language and the world?
- How can we build language technologies that support diversity within and across languages?

Language Change

- Language is constantly changing through its use
- Individual language users constantly adapt language (and general behavior) to feedback interlocutors provide
- We are constantly innovating language use, e.g., by proposing new words to things
- How do individuals contribute to language change?
- When are proposed changes taken up by a language community?
- How do the intrinsic dynamics of language challenge our current assumptions about language modeling?

Information

- Thinking about language purely as data sampled from a statistical process
- Compared to other kinds of data (e.g., images), there are fundamental differences in its structure: e.g., it's highly discontinuous
- What fundamental properties does language hold with respect to information theoretic principles?
- How do these principles guide the form language takes?
- Where do uncertainties lie in our utterances, and why?
- How do these properties challenge assumptions made by current language models?

Thursday, January 18
Meeting

- Todos before next meeting
  - Fill the intro questionnaire
  - Check your role assignment (will post the sheet this evening)
  - Read paper for Thursday
  - Role representatives: prep ahead of time if necessary

- **Reading:** Beckner et al. 2009, Language is a complex adaptive system