Class Info/Logistics

- COMP 790.139 ‘Language Grounding for Robotics’
- **3 UNITS**
- Time: Tuesdays 9.30am-12.00pm
- Room: FB008
- Office Hours: Tue 12.00-1.00pm (by appointment), SN258
- **Course Email**: nlpcomp790unc@gmail.com
About Me

- Asst. Professor, CS, UNC (joined Fall 2016)
- PhD, UC Berkeley, 2008-2013

Research Interests:
- Past: Syntactic parsing, coreference resolution, taxonomy induction, world knowledge and commonsense induction
- Current: Multimodal and embodied semantics (i.e., language with vision and speech, for robotics); human-like language generation and Q&A/dialogue; interpretable and structured deep learning

- Office SN258
- Webpage: http://www.cs.unc.edu/~mbansal/, Email: mbansal@cs.unc.edu
Your Introductions

Please say your:

- Name
- Department/degree/major
- Research interests (ML/AI/NLP/CV experience? Coding?)
- Fun fact 😊

(send me a few lines of intro at nlpcomp790unc@gmail.com)
About the Course (and its Goals)

- Research-oriented seminar course! We will read lots of interesting papers, brainstorm, and do fun novel projects!

- Basic NLP and machine learning (some deep learning) experience expected – check fall2016 NLP intro slides!

- We’ll cover several latest research topics in NLP+robotics

- You will read papers in advance and write summaries

- You will present papers in class

- Brainstorm regularly and code + write up fun/novel projects!

- Some lecture(s) on academic/research quality paper writing
Prerequisites

- Graduate research-level class
- Some machine learning and coding experience is definitely expected!
- Moreover, some NLP and RL background is highly recommended
- Projects will require research and coding skills; summaries and paper presentation will require solid ML clarity
Expectations/Grading

- Paper presentation (15%)
- Paper written summaries (25%)
- Class participation, discussion and brainstorming (20%)
- Project reports and presentations (40%) (midterm = 15% and final = 25%)
Lateness Policy

- Written summaries are due *before class* by email to nlpcomp790unc@gmail.com

- First paper summary submission will have no late penalties.

- After that, for every week's summary submission, there will be a 25% value reduction per late day.

- Other lateness policies (for projects, etc.) will be sent via email during the semester
Collaboration Policy

- Paper summaries have to be written and submitted individually.
- Projects are encouraged to be done in pairs, with clearly outlined contributions from each team member.
- But individual projects are fine too (e.g., if it relates to your current research).
Paper Presentation

- Lead discussion for 1-2 papers on a topic some week (may be done in pairs/groups depending on class size)

- Read related papers and present background to audience

- Present task and model details of given paper

- Present demo’s of related code, etc.

- Ask interesting questions to initiate brainstorming

- Mention some next steps, future work, extension ideas!
Paper Written Summaries

- 0.5-1 page (per paper) write-up for every week’s paper(s)
- Describe the task
- Summarize the method
- Explain the novelty
- Discuss the next steps or potential improvements
Class Participation and Brainstorming

- Audience students expected to take part in lively discussion every paper reading!

- After every topic gets completed (i.e., several papers in 2-3 weeks), we will have a brainstorming and ‘idea-generation’ session!

- Exact details to be announced soon but students expected to submit and discuss novel idea(s) on the whole general topic, e.g., new related task or dataset, new approach to existing task, combinations of tasks/approaches, etc.

- Don’t hesitate to propose fancy ideas 😊, but try to keep them grounded/feasible and think of how to approach them realistically (in terms of datasets, models, speed, memory, etc.)
Project

- This is a ‘Reading, Coding, and Writing’ class!
- Students will pick (early) their favorite topic among latest cutting-edge research topics covered in class
- And will try a novel idea (implementing + extending or original) -- I am happy to discuss details!
- Midterm and final report + presentation (and possibly some updates)
- Might be in pairs/groups depending on final class size
- Use conference style files and aim for conference-quality papers
- Will have some lecture(s) on research-quality paper writing
Reference Books


- **SLP3**: Some draft chapters of the third edition are available online at [https://web.stanford.edu/~jurafsky/slp3/](https://web.stanford.edu/~jurafsky/slp3/)


- **ML Background**: Andrew Ng’s Coursera Machine Learning course [https://www.coursera.org/learn/machine-learning](https://www.coursera.org/learn/machine-learning)

Course Syllabus/Topics (tentative)

- interpreting and executing verbal instructions for navigation, articulation, manipulation, assembly, skill learning, etc.

- human-robot collaboration and dialogue for learning new subactions, mediating shared perceptual basis, referring expression generation, etc.

- grounding and language learning via dialogue-based and interactive games.

- automatic language generation for embodied tasks.

- grounded reinforcement learning.

- grounded knowledge representations (mapping language to world).

- machine learning models (structured and deep), datasets, and metrics for embodied language.
What is NLP?

- Question answering
What is NLP?

Question answering

Google search result for "What was the U.S. population when Bernie Sanders was born?"

About 1,620,000 results (0.67 seconds)

United States of America / Population (1941)

133.4 million
1941
What is NLP?

Question answering

Google search for "which countries border the black sea"

This major inland sea is bordered by six countries — Romania and Bulgaria to the west; Ukraine, Russia, and Georgia to the north and east; and Turkey to the south. Additionally, it is impacted by another 10 nations through the five major rivers that empty into the Black Sea, the largest of which is the Danube River.

Black Sea Geography - College of Earth, Ocean, and Environment
https://www.ceoe.udel.edu/blacksea/geography/index.html

University of Delaware
What is NLP?

- Machine Translation

This is an example of machine translation.

यह मशीन अनुवाद का एक उदाहरण है

Yaha maśīna anuvāda kā ēka udāharaṇa hai
What is NLP?

Sentiment Analysis

**Sentiment Analysis with Python NLTK Text Classification**

This is a demonstration of sentiment analysis using a NLTK 2.0.4 powered text classification process. It can tell you whether it thinks the text you enter below expresses positive sentiment, negative sentiment, or if it's neutral. Using hierarchical classification, neutrality is determined first, and sentiment polarity is determined second, but only if the text is not neutral.

---

### Analyze Sentiment

**Language**

| english |

**Enter text**

It always amazes me how Universal never cares to create anything remotely clever when it comes to their animations, and so once again they come up with a harmless little story that wants to be cute and funny (which it is sometimes) but is only bound to be quickly forgotten.

---

### Sentiment Analysis Results

The text is **neg**.

The final sentiment is determined by looking at the classification probabilities below.

**Subjectivity**

- neutral: 0.3
- polar: 0.7

**Polarity**

- pos: 0.2
- neg: 0.8
What is NLP?

- Lohan charged with theft of $2,500 necklace
- Plead not guilty
- Judge set bail at $40,000
- To reappear in court on Feb 23
What is NLP?

- Natural Language Generation: Conversation/Dialogue

context
because of your game?

message
yeah i’m on my way now

good luck!

Because of your game? Yeah I’m on my way now. Good luck!
What is NLP?

Natural Language Generation: Image Captioning

Results

Tags

- authors
- scones
- luncheon
- breakfast
- seder

Nearest Caption in the Training Dataset

a man cuts a cake while children sit around at the table, looking on.

Generated Captions

- two people at a table with a cake.
- the two people are having a meal that is in a party.
- a man and two children in a blue table with a cake.
- a man sitting at a table with a bunch of cake on it.
- a man and woman sitting at a table with cake at a party.

back
What is NLP?

Natural Language Generation: Visual Question Answering

- What color are her eyes?
  - What is the mustache made of?

- How many slices of pizza are there?
  - Is this a vegetarian pizza?

- Is this person expecting company?
  - What is just under the tree?

- Does it appear to be rainy?
  - Does this person have 20/20 vision?
What is NLP?

- Automatic Speech Recognition
What is NLP?

- Task-based instructions, e.g., navigation, grasping, manipulation, skill learning

Turn right at the butterfly painting, then go to the end of the hall
What is NLP?

- Task-based instructions, e.g., navigation, grasping, manipulation, skill learning

"Cut some onions, and add to broth, stir it"
Some Sample NLP+Robotics Papers

- Navigation Instruction Following

Learning to interpret natural language navigation instructions from observations. Chen and Mooney. AAAI 2011.

Weakly supervised learning of semantic parsers for mapping instructions to actions. Artzi and Zettlemoyer. TACL 2013.

Navigation + Manipulation Instructions

Figure 1: A target robotic platform for mobile manipulation and navigation (Teller et al., 2010), and sample commands from the domain, created by untrained human annotators. Our system can successfully follow these commands.

<table>
<thead>
<tr>
<th>Command from the corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Go to the first crate on the left and pick it up.</td>
</tr>
<tr>
<td>- Pick up the pallet of boxes in the middle and place them on the trailer to the left.</td>
</tr>
<tr>
<td>- Go forward and drop the pallets to the right of the first set of tires.</td>
</tr>
<tr>
<td>- Pick up the tire pallet off the truck and set it down</td>
</tr>
</tbody>
</table>

(a) Robotic forklift                                                                 | (b) Sample commands                                                                 |

Figure 4: A sequence of the actions that the forklift takes in response to the command, “Put the tire pallet on the truck.” (a) The search grounds objects and places in the world based on their initial positions. (b) The forklift executes the first action, picking up the pallet. (c) The forklift puts the pallet on the trailer.

Navigation Dialogue

Fig. 1. A user gives a tour to a robotic wheelchair designed to assist residents in a long-term care facility. (Left) The guide provides an ambiguous description of the kitchen’s location. (Right) When the robot is near one of the likely locations, it asks the guide a question to resolve the ambiguity.
Navigation Instruction Generation

Fig. 4. Participants’ field of view in the virtual world used for the human navigation experiments.

Fig. 8. Examples of paths from the SAIL corpus that ten participants (five for each map) followed according to instructions generated by humans and by our method. Paths in red are those traversed according to human-generated instructions, while paths in green were executed according to our instructions. Circles with an “S” and “G” denote the start and goal locations, respectively.
Some Sample NLP+Robotics Papers

Assembly Instruction Execution

<table>
<thead>
<tr>
<th>Scene</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Scene Image]</td>
<td>Move the block that is currently located closest to the top left corner to the bottom left of the table, slightly higher than the block in the bottom right corner.</td>
</tr>
<tr>
<td><strong>Error:</strong></td>
<td></td>
</tr>
<tr>
<td>![Scene Image]</td>
<td>Move the block closest to the top left corner so it is above half a block length to the right of the blocks near the lower left corner of the table.</td>
</tr>
</tbody>
</table>

**Table 6:** Above are two commands and the worlds they apply to. Below we see the prediction error of our best model.
Some Sample NLP+Robotics Papers

Recipes: Bakebot (http://projects.csail.mit.edu/video/research/robo/bakebot_final.mp4)

Fig. 3 Architecture of the BakeBot system. The NL system processes the plain text recipe, producing a high-level plan which is sent to the robot. For each instruction in the high-level plan, the motion planner assembles a motion plan and executes it on the PR2 robot.
Fig. 1. **Natural Language Instructions to sequence of instructions** for a given new environment. Our approach takes description in natural language and sequences together robotic instructions that are appropriate for a given environment and task. Note that the NL instructions are often ambiguous, and are incomplete, and need to be grounded into the environment.
Some Sample NLP+Robotics Papers

Recipes: Tell Me Dave (http://tellmedave.cs.cornell.edu/)

Fig. 4. **Robot Experiment.** Given the language instruction for making the dessert ‘Affogato’: ‘Take some coffee in a cup. Add ice cream of your choice. Finally, add raspberry syrup to the mixture.’, our algorithm outputs a sequence that the PR2 executes to make the dessert. (Please see the video.)
Recipes: RoboBarista (http://robobarista.cs.cornell.edu/)

Fig. 5. Screen-shot of Robobarista, the crowd-sourcing platform running on Chrome browser. We have built Robobarista platform for collecting a large number of crowd demonstrations for teaching the robot.
Learning New Actions by Situated Human-Robot Dialogue

Figure 1: An example setup and dialogue. Objects are marked with labels only for the illustration purpose.

H1: Stack the blue block on the red block on your right.
R1: How to stack, could you teach me?
H2: Grab the blue block.
R2: What do you mean by grab?
H3: Open your gripper.
R3: Ok.
H4: Move to the blue block.
R4: Sure.
H5: Close gripper.
R5: Alright then.
H6: Now you achieve the grab action.
R6: Ok, got it.
H7: Move the blue block to the top of the right red block.
R7: Understood.
H8: Open gripper.
R8: Sure.
H9: Ok, now you are done with the stack action.
R9: Ok.
and executability of actions, and representing and reasoning with the preconditions, effects, et al. symbolic planner formalized using action language typing in natural language. The agent generated queries to interacted with the agent through a graphical user interface by 6.1 Implementationing System (ROS) (Segbot) as shown in Figure 5 (Left) using the Robot Operating

The agent was integrated into a Segway-based robot platform in addition to learning misspelling corrections and new referring Turk dialogs where the correct goal was reached. Means in underlined bold differ significantly (Figure 3: Learning To Interpret Language Commands via Dialogue

By matching some Sample NLP+Robotics Papers

Learning To Interpret Language Commands via Dialogue

Some Sample NLP+Robotics Papers

Learning To Interpret Language Commands via Dialogue

Figure 3: The Mechanical Turk interface for the delivery task. This abridged conversation is from a Turker in training batch 0, when the system had access to only the seed lexicon. Because of this conversation, the agent learned that “calendar” and “planner” mean “calendar” during retraining.

Figure 5: Left: Robot platform (Segbot) used in experiments. Right: Segbot architecture, implemented using Robot Operating System (ROS).
Abstract

We introduce a new language learning setting relevant to building adaptive natural language interfaces. It is inspired by Wittgenstein’s language games: a human wishes to accomplish some task (e.g., achieving a certain configuration of blocks), but can only communicate with a computer, who performs the actual actions (e.g., removing all red blocks). The computer initially knows nothing about language and therefore must learn it from scratch through interaction, while the human adapts to the computer’s capabilities. We created a game called SHRDLURN in a blocks world and collected interactions from 100 people playing it. First, we analyze the humans’ strategies, showing that using compositionality and avoiding synonyms correlates positively with task performance. Second, we compare computer strategies, showing that modeling pragmatics on a semantic parsing model accelerates learning for more strategic players.

1 Introduction

Wittgenstein (1953) famously said that language derives its meaning from use, and introduced the concept of language games to illustrate the fluidity and purpose-orientedness of language. He described how a builder B and an assistant A can use a primitive language consisting of four words—‘block’, ‘pillar’, ‘slab’, ‘beam’—to successfully communicate what block to pass from A to B. This is only one such language; many others would also work for accomplishing the cooperative goal. This paper operationalizes and explores the idea of language games in a learning setting, which we call Interactive Learning through Language Games (ILLG). In the ILLG setting, the two parties do not initially speak a common language, but nonetheless need to collaboratively accomplish a goal. Specifically, we created a game called SHRDLURN, in homage to the seminal work of Winograd (1972). As shown in Figure 1, the objective is to transform a start state into a goal state, but the only action the human can take is entering an utterance. The computer parses the utterance and produces a ranked list of possible interpretations according to its current model. The human scrolls through the list and chooses the intended one, simultaneously advancing the state of the blocks and providing feedback to the computer. Both the human and the computer wish to reach the goal state.

Figure 1: The SHRDLURN game: the objective is to transform the start state into the goal state. The human types in an utterance, and the computer (which does not know the goal state) tries to interpret the utterance and perform the corresponding action. The computer initially knows nothing about the language, but through the human’s feedback, learns the human’s language while making progress towards the game goal.
Since we are addressing referential communication to mediate perceptual differences between humans and robots, we intentionally applied only a simple computer vision algorithm [19]. Here, except for Object 4 which is correctly recognized as a banana with a confidence of 0.65, the rest of objects are mis-recognized. The numerical values related to color, location, and the size of bounding boxes are also captured by the internal representation. Figure 1(c) shows an example of the embodied collaborative referential process for the robot to gradually lead the human to the target object (i.e., the red apple as shown by the arrow in Figure 1(a)). Note the robot takes the human's perspective when describing the objects in the scene.

Incorporation of robot’s gesture in referring acts. Pointing gestures from a robot are combined with verbal descriptions to generate referring expressions. Here, we treat the cost of gesture generation as a feature and incorporate it with other features into the collaborative models [7]. Since it is expensive to conduct a large scale in-lab study to learn feature weights from real-time human-robot interaction, we directly adopt the features and their learned weights from our web-based study [7] as shown in Table 1 (Feature 1-22). We then explicitly add the cost of the gesture generation as an additional feature and set its weight to 0 (Feature 23).

The cost of a pointing gesture depends on several factors: the distance from the robot to the target object, the size of the target object, adjacency of other objects to the target object, etc. Inspired by previous work on the costs of pointing, [10] and [21], we define the cost of a pointing gesture to an object as follows:

\[
\text{cost} = \begin{cases} 
\log_2 \left(1 + \frac{1}{\text{Distance} \cdot \text{Size}}\right) & \text{if the object to be described is in a group} \\
\log_2 \left(1 + \frac{1}{\text{Distance} \cdot \text{Size}} + 1\right) & \text{otherwise} 
\end{cases}
\]

Given a situation \(s\), the robot will apply the features and their associated weights in Table 1 to calculate \(Q(s, a)\) and then choose the generation action \(a^*\) that maximizes \(Q(s, a)\).

Note that incorporating gesture does not necessitate that the generated referring expressions will always include gestures. Whether gesture is used or not depends on how it weighs against other features extracted from the environment. For example, in Figure 1(c), \(R_1\) includes the pointing gesture as part of the referring expression, while \(R_2\) does not.

Incorporation of eye gaze as feedback. Previous psycholinguistic studies have shown that human eye gaze directly links with language comprehension [27]. Immediately after hearing a referring expression, the hearer’s eyes move to the objects being referred to. Motivated by this finding we incorporate the user’s real-time gaze and verbal information as intermediate feedback in the installment model.

To incorporate human gaze feedback, we must find how the gaze is distributed among each of the objects in the scene over the time immediately following an RE such as in Figure 1(c). Specifically, starting from the onset of an RE uttered by the robot, we capture at least 200 gaze readings (taking on average 7.7 sec) from the participant and calculate a distribution showing which objects the gaze is drawn to most often (i.e., which objects draw the most gaze readings). Based on this distribution, we applied a simple criterion to identify the focused object: the object with the highest number of readings where its reading number is at least twice as high as the reading number of the second highest object.

The gaze feedback can be incorporated into the installment model with or without verbal feedback.

- **Gaze Only Feedback**
  - If a focused object can be identified based on the gaze, then the installment is considered successful and the focused object becomes the target.

- **Verbal Only Feedback**
  - If a focused object cannot be identified based on the gaze, then the installment is considered unsuccessful and the system asks for additional verbal input.

- **Mixed Feedback**
  - If both gaze and verbal feedback are available, the system uses this information to make a decision about the installment.

Next Week

- Website will be updated with initial paper list very soon
- Once paper list is up, start volunteering for paper presentation!
- We might do pairs/groups of students depending on class size
- Project details and dates will be announced soon