CS224v

Conversational Virtual Assistants with Deep Learning

Lecture 2: Project Introduction

Monica Lam
Conversational Virtual Assistant

Music
- Spotify

Podcasts
- Spotify

Radio
- TuneIn

News
- SmartNews

Restaurants
- Yelp

Search
- Bing

Weather
- Yr.no

Jokes
- Dad Jokes

Reminder

Timer

Internet of Things

Appliances
- Thermostat
- Switch
- Light
- Fan
- Door
- Lock
- Window Cover
- Vacuum Cleaner

Sensors
- Temperature
- Motion
- Illuminane
- Humidity
- Flood
- Battery
- Ultra-violet
Open-Source, Private Assistant

Demo on https://oval.cs.stanford.edu

1. On the web: genie.stanford.edu

2. Baidu
   smartspeaker development kit

3. Home Assistant
   Open-source Home Automation Project
   - 100,000+ users
   - A local, private gateway controlling over 1000 IoT devices
   - Genie distributed as voice interface

Other partners: Picovoice, SmartNews, Yelp
Our Social Intelligent Agent Research

Motivation: Knowledge Navigator, Seniors’ Companion

Integrating together:

• Assistance: provide correct answers & actions  \(\rightarrow\) Semantic Parsing
• Goal-directed conversation  \(\rightarrow\) Dialogue Flow
• Chitchat conversation
• Common knowledge
• Social intelligence
• Multimodal interaction

as a high-level dialogue scripting language: GenieScript
Current Research Projects

1. Do LLMs have Social Intelligence?

2. Integrating LLMs, Dialogue Trees, Semantic Parsing
   a. High-Level Dialogue Trees: Multi-lingual Multi-Domain Task-Oriented Agents in a Week
   b. LLM-Based Semantic Parsing
   c. Wiki-LLM: Adding Factuality to LLM

3. Socially Intelligent Assistants
1\textsuperscript{st} Investigation
Autism Spectrum Disorder (ASD)

- ASD
  - Lacks social intelligence
  - Restricted and repetitive behaviors or interest
- 1 in 44 US children diagnosed with autism by age 8.
- High unemployment rates
- Lack friendships and romantic relationships

Mason Wang, Ethan Chang, Stefanos Polychronopoulos (high school students)
Sina Semnani (CS PhD), Lynn Koegel (Stanford Psychiatry), Monica Lam
Teaching Social Skills

- Dr. Koegel (Psychiatry, Stanford)
  - taught adults (ASD, social anxiety) social skills through repetitive training
  - Autistic people cannot transfer learning
  - Too few trained clinicians
    → automation makes training affordable

- Do LLMs have social intelligence?
  Can they learn social intelligence with training?
Example of an Empathy Exercise

I went to see a movie last night, I fell and broke my ankle.

What movie did you see?

Not quite. You friend was hurt, you should first show that you are concerned. A better reply could be “Oh no, is it hurting bad?”
Improve your social conversation using AI.

Noora is a chatbot who guides you through a diverse set of social scenarios.

Developed by Stanford's Open Virtual Assistant Lab

Good reply.

Thanks for showing happiness for me and asking about my experience.
Traditional: Fine-tune with Training Data

1. Ask crowdworkers for possible responses to pre-defined scenarios
2. Manually annotate the classification and explanation
3. Fine-tune pre-trained language models with many samples

LLM: Prompt with Few-Shot Examples

1. Use LLM to generate representative responses to a few scenarios
2. Use LLM to classify and explain
3. Handpick a few representative examples
Use GPT-3 to Generate Different Replies

- GPT-3 knows different ways people may reply
- Use a latent variable (attitude) to coax it into generating them

List ten different interesting ways you can reply and then give a corresponding reply for each approach. I say, “I just finished a good book!”

2 few-shot examples

(1) Self-centered: “Reading is one of my hobbies.”
(2) Off-topic: “What time is it?”

GPT-3 output

(3) Supportive: "What's the book about? I'm so happy for you that you enjoyed it!"
(4) Controlling: "You should read more books! You'll get smarter."
(5) Jealous: "I don't have time to read. I'm too busy."

...
Prompting GPT-3 (No fine-tuning)

1st 5 example replies
You said, "I am going to the grocery store."
(1) I replied, "Great! I hope you find everything you need."
Feedback: Good reply. Your reply is happy and encouraging.
...

2nd 5 example replies
You said, "This is my cubicle."
(1) I replied, "You have an ugly cubicle. Why are you showing me?"
Feedback: Bad reply. I feel hurt by your reply because it is mean. You could ...
...

new statement + reply pair
You said, "Tomorrow is supposed to be warm."
(1) I replied, "Geez, I hope I don't melt!"
Feedback: Good reply. You sound like you're joking around and that you're excited for the weather.

“good” humor!
Prompt-Based Generation

- GPT-3 has social intelligence
  - Understands social norms, gives expected responses
- Explanations to user input are decent
- Sometimes gets it wrong (not critical for this app)
- Research starts with studying the hard cases
- Fine-tuning may cause problem for overfitting (Hard topic to have good coverage)
Ongoing Research

- Broaden the study of social intelligence
- Anti-toxicity, prosocial behavior, extract persona, customize response based on persona
- Exploring “thoughtfulness”
  - GPT-3 generates 3 responses, then reasons, then pick.
- Noora was well received in a preliminary study; will finish and do a real study with users (with Dr. Koegel)
LLMs: Pros and Cons

• Pros: Plausible responses with fluency and embellishment
  Demonstrate common knowledge, social intelligence

• Cons:
  1. Hallucination, no understanding of semantics, not grounded with external resources
     • Facts from live knowledge bases or APIs
     • Performing actions (e.g. turn off the light)
  2. No control over low-level expressions (including human values)
  3. No control over high-level choices: purpose, long term memory

Cannot use as is!
Current Research Projects

1. Do LLMs have Social Intelligence?
2. **Integrating Dialogue Trees, Semantic Parsing, LLMs**
   a. High-Level Dialogue Trees: Robust Multi-lingual Multi-Domain Task-Oriented Agents in a Week
   b. LLM-Based Semantic Parsing
   c. Wiki-LLM: Adding Factuality to LLM
3. Socially Intelligent Assistants
Prefrontal Cortex Inspired Architecture
(Dialogue Tree, Semantic Parsing, LLMs)

Programmable Executive Controller

User NL

Agent NL

Large Language Model (LLM)

Semantic Parser

NL

External DB/API

TT
A Restaurant Assistant

Request
I’d like to find a moderately priced restaurant

ProposeOne
I have Terun. It’s a moderately priced restaurant that serves pizza.

ProposeN
I found Terun and Coconuts. Both are moderately priced.

AskAction
I like that. Can you help me book it? I need it for 3 people.

SearchRefine
I don’t like pizza. Do you have something Caribbean?

InfoQuestion
Can you tell me the address of Terun?

Large Language Models

AutoQA: From Databases To Q&A Semantic Parsers With Only Synthetic Training Data
Silei Xu*, Sina J. Semnani*, Giovanni Campagna, Monica S. Lam
Domain-Independent State Machines Tested on MultiWOZ

Restaurant  Hotel  Taxi  Train  Attraction

Manual Annotations: 2% of original
Turn-by-turn accuracy: 80% → State-of-the-art

A Few-Shot Semantic Parser for Wizard-of-Oz Dialogues with the Precise ThingTalk Representation
Giovanni Campagna, Sina J. Sennani, Ryan Kearns, Lucas Jun Koba Sato, Monica S. Lam
In Findings of ACL, 2022
• 1st high-level programming language for dialogue agents
  • Neural: a contextual semantic parser for conversations
  • Sample efficient with synthesized multi-lingual conversational data
  • Supports multi-lingual user-defined dialogue flow
  • A formal conversational semantics
  • To handle ellipses and co-references
  • Supports mixed-agent initiatives

Shicheng (George) Liu, Jackie Yang, Monica Lam: To submit to PLDI 2023
No-Code Mixed-Initiative Agent

User

I want to transfer money

What account would you like to transfer money to?

12512032

Got it. How much money would you like to transfer?

How much money do I have in my account?

You have $1,050

How much money would you like to transfer?
New Arch → Many Research Projects

Programmable Executive Controller

User NL → Large Language Model (LLM) → Semantic Parser

Agent NL

External DB/API

TT
Current Research Projects

1. Do LLMs have Social Intelligence?
2. Integrating Dialogue Trees, Semantic Parsing, LLMs
   a. High-Level Dialogue Trees: Robust Multi-lingual Multi-Domain Task-Oriented Agents in a Week
   b. LLM-Based Semantic Parsing
   c. Wiki-LLM: Adding Factuality to LLM
3. Socially Intelligent Assistants
Robust Multi-lingual Multi-Domain Task-Oriented Agents in \texttt{geniescript} in a Fortnight

Hypothesis

- It takes less than 20 lines of code per domain
- Automatic training with few-shot + synthesized data
  - Performance within 10\% of training with full data
- Write output strings in a new language → agent in a new language
  - Performance within 10\% of original language

Mehrad Moradshahi, Shicheng (George) Liu, Jackie Yang, Monica Lam
Multi-Lingual Question Answering with Local Entities in 1 Day

Genie
Train parser with auto-translated data + 2% manually translation

SOTA
On-the-fly translation

Localizing Open-Ontology QA Semantic Parsers in a Day Using Machine Translation
Mehrad Moradshahi, Giovanni Campagna, Sina J. Semnani, Silei Xu, Monica S. Lam
Experiment to Run

- RiSAWOZ: 12 domains
  largest & best-annotated multi-domain task-oriented WOZ dataset in Chinese (Tianjin U)

- Compare few-shot with full-training

- Our current X-RiSAWOZ project: few-shot data set for
  
  - Chinese, English (Tianjin U), French (CEA), Hindi (IIIT), Hindi-English (Microsoft), Korean (Han Yang U).

- Compare accuracies across languages
# New Kinds of Assistants

## Example: A Workout Assistant

<table>
<thead>
<tr>
<th>USER</th>
<th>ASSISTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;I want to create a work day morning routine&quot;</td>
<td>How can I help you?</td>
</tr>
<tr>
<td>&quot;Set a timer for my planks at 2 minutes, 3.5 minutes, 5 minutes, and 6.5 minutes&quot;</td>
<td>OK what would you like for step 1.</td>
</tr>
<tr>
<td>&quot;Read me emails from my boss&quot;</td>
<td>Got it, what is step 2.</td>
</tr>
<tr>
<td>&quot;Read me daily digest news from CNN&quot;</td>
<td>Got it, what is step 3.</td>
</tr>
<tr>
<td>&quot;That's it&quot;.</td>
<td>Got it, what is step 4.</td>
</tr>
<tr>
<td></td>
<td>OK, for your morning routine, I will …</td>
</tr>
</tbody>
</table>
Workout Assistant in Action

Start my morning routine.

Your timer is set at 2 minutes, 3.5 minutes, 5 minutes, and 6.5 minutes. Here is your email from your boss: ....

[after one email]: Please forward this email to my secretary.

Ok. Got it, here is the next one: ...

[On a long message]: Skip

Here is the CNN news ...

[Interrupting:] Where is ‘Machu Picchu’?
Current Research Projects

1. Do LLMs have Social Intelligence?
2. Integrating Dialogue Trees, Semantic Parsing, LLMs
   a. High-Level Dialogue Trees: Robust Multi-lingual Multi-Domain Task-Oriented Agents in a Week
   b. LLM-Based Semantic Parsing
   c. Wiki-LLM: Adding Factuality to LLM
3. Socially Intelligent Assistants
Semantic Parsing

Fine-tuned on Bart. Few-shot training data + Synthesized data

### Examples of Long-Tail Questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Alexa</th>
<th>Google</th>
<th>Siri</th>
<th>Genie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show me restaurants rated at least 4 stars with at least 100 reviews</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Show restaurants in San Francisco rated higher than 4.5</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>What is the highest rated Chinese restaurant near Stanford?</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>How far is the closest 4 star restaurant?</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Who works for W3C and went to Oxford?</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Who worked for Google and lives in Palo Alto?</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Who graduated from Stanford and won a Nobel prize?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Who worked for at least 3 companies?</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Show me hotels with checkout time later than 12PM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Which hotel has a pool in this area?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Long-tail Restaurant Questions

AutoQA: From Databases To Q&A Semantic Parsers With Only Synthetic Training Data
Silei Xu*, Sina J. Semnani*, Giovanni Campagna, Monica S. Lam
## Challenge: Common Sense/Knowledge

**FACT:** King Charles III’s stepchildren are Tom Parker Bowles and Laura Lopes

WikiData: does not have “stepchildren”

<table>
<thead>
<tr>
<th>Who are King Charles III’s stepchildren?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The stepchildren of King Charles III are Princess Louise of Belgium and Princess Stéphanie of Belgium.</td>
</tr>
<tr>
<td>(Both children of Leopold II of Belgium)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who are Prince Charles’s stepchildren?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Prince Charles's stepchildren are Sarah, Jane, and Andrew.</td>
</tr>
</tbody>
</table>
Chain-of-Thought Prompting
FACT: King Charles III’s stepchildren are Tom Parker Bowles and Laura Lopes

1. **What are stepchildren?**
   *Stepchildren are* children of a spouse with another partner.

2. **Stepchildren are children of a spouse with another partner.** King Charles III’s stepchildren are Princess Louise of Belgium and Princess Stéphanie of Belgium. *(Both children of Leopold II of Belgium)*

3. **Stepchildren are children of a spouse with another partner.** Prince Charles’s stepchildren are Camilla Parker Bowles’s children, Laura Lopes, Tom Parker Bowles, and Laura Parker Bowles.

*We need to combine LLMs with Semantic Parsing*
LLMs for Semantic Parsing Projects

1. Wikidata schemas: too large (10,278 properties)
   - Previous: Fine-tuned BART with few-shot + synthesized data
   - Fine-tune LLM to generate formal queries
   - Experiment with more sample-efficient training

2. Wikidata schemas: too unstructured to remember all properties
   - Fetch relevant schemas on-the-fly

3. Complex questions
   - Use chain-of-thought prompting to include facts up front

Silei Xu, Sina Semnani, Mehrad Moradshahi, Monica Lam
Current Research Projects

1. Do LLMs have Social Intelligence?

2. Integrating Dialogue Trees, Semantic Parsing, LLMs
   a. High-Level Dialogue Trees:
      Robust Multi-lingual Multi-Domain Task-Oriented Agents in a Week
   b. LLM-Based Semantic Parsing
   c. Wiki-LLM: Adding Factuality to LLM

3. Socially Intelligent Assistants
Wiki-Genie

A conversational chatbot on Wikipedia website

**User**

Who wrote Jurassic Park?  
That was such a good movie.  
What was his first big movie?  
And the second?  
Westworld? I didn’t know that.  
Who starred in it?

**Wiki-Genie**

Michael Crichton  
Andromeda Strain in 1971  
Westworld in 1973  
Yul Brynner, Richard Benjamin
A true resource: Not just an academic exercise

- Our semantic parser for single commands is at about 30-40% on QALD7 (state of the art)
- Wikidata is a global crowdsource resource with over 22,000 contributors
- Key: Make Wiki-Genie self-learning get continuous feedback and correction through multimodal conversations

Sina Semnani, Jackie Yang, Silei Xu, Shicheng Liu, Monica Lam
User: Do you know about Queen Elizabeth II?
I was just reading an article that mentioned her name.
GPT-3: Yes, I know a lot about her!
User: Can you tell me?
GPT-3: Queen Elizabeth II is the current monarch of the United Kingdom. She has been on the throne since 1952 and is the longest-reigning British monarch.
Adding Thoughtfulness to GPT-3

Loop
- Ask GPT-3 to generate response
- Ask GPT-3 to question itself
- Consult Wiki
- If not fact-checked Add facts as prompt

**Output:** Queen Elizabeth II is the current monarch of the United Kingdom. She has been on the throne since 1952 and is the longest-reigning British monarch.

**Questions:** Is Queen Elizabeth II the current monarch of the United Kingdom? Did Queen Elizabeth II start her reign in 1952? …

**Wiki-Genie:** No, Yes, …

**Prompt:** Queen Elizabeth II is not the current of the United Kingdom. Queen Elizabeth II started her reign in 1952... Can you revise the previous statement to take this new information into account?

**Output:** Queen Elizabeth II was the monarch of the United Kingdom until her death in 2020. …

**Questions:** When did Queen Elizabeth II pass away?

**Wiki-Genie:** 2022
Current Research Projects

1. Do LLMs have Social Intelligence?
2. Integrating Dialogue Trees, Semantic Parsing, LLMs
   a. High-Level Dialogue Trees: Robust Multi-lingual Multi-Domain Task-Oriented Agents in a Week
   b. LLM-Based Semantic Parsing
   c. Wiki-LLM: Adding Factuality to LLM
3. Socially Intelligent Assistants
Social Intelligent Assistants

- Previous architecture (Chirpy, Grillbot)
  - Multiple modules (LLM, Question Answering)
- Fine-tune 1 LLM to generate
  - A semantic parse for questions or actions
  - Free-text answers
    - Add thoughtfulness/fact-checking to revise answers
- Experiments: a TV, movie, sports chatbot
  - Can be performed in parallel with above projects
- Next step
  - Extract personas from conversation; Adapt to personas
Examples of Previous Course Projects

• **A Multimodal Building Management Assistant** (Junwen Zheng)
  • Brings digitization to the “hardhat” community

• **RUSS: Rapid Universal Support Service**
  (Nancy Xu, Nancy Xu, Sam Masling, Michael Du)
  • Train once and answer calls for different companies
Facility Management MultiModal Assistant
Junwen Zheng, CIFE, Q4 2021, Q1 2022

Construction industry: 2\textsuperscript{nd} least digitized industry
Digital information inaccessible to blue collar workers
• Not tech savvy
• Hands are busy
• On the job, not at the desk

Voice interface: access info on the job naturally
Built with 100 manually annotated commands + synthesized data
for Facilities Management
Automated Call Agents

Research Question:
Can we train one universal agent that reads instructions on online help centers?

Create an Account
1. Select the Account button in the upper-right corner of Walmart.com
2. Click Create an Account
3. Enter your name and email address on the sign-up screen
4. Create a password and confirm it
5. Select Create Account

Redeem a Gift Card
1. Find the claim code.
2. Go to Redeem a Gift Card.
3. Enter your claim code and select Apply to Your Balance.

Find your AD Account
0. Click Ads to open Ads Manager
0. Click Reporting
0. Click the Ads tab under the graph
0. Scroll down to the list of ads
0. The 12-digit number beginning with 687 under the ad name is the ad ID
RUSS
Rapid Universal Support Service
Nancy Xu, Sam Masling, Michael Du, Q4, 2020

- Trained once with synthesized data
- Extended ThingTalk for web operations
- Experiment
  - 22 different online help centers
  - 80 tasks, 741 instructions
  - Overall accuracy: 77%
  - Semantic parsing accuracy: 87%
  - Users prefer RUSS over the web

Grounding Open-Domain Instructions to Automate Web Support Tasks
Nancy Xu, Sam Masling, Michael Du, Giovanni Campagna, Larry Heck, James Landay, Monica S Lam