CS 294S/294W
Building the Best Virtual Assistant
A Research Project Course

Monica Lam
Stanford University
lam@cs.stanford.edu

Supported by NSF Grant #1900638
Why a Remote Research Course?

A welcomed change from Zoom lectures.
Expose students to the exciting world of research.
## Virtual Assistants!

| A once-in-20-years research opportunity | Mainframe, PCs, web, mobile/ubiquitous
|                                       | Entire web available by voice in all languages |
| Vision                                 | 23M voice interface developers |
| New technical approach                 | Annotating real data $\rightarrow$ training-data engineering |
| A new NLP data engineering tool chain  | Virtual assistant programming language
|                                       | Grammar-driven data synthesis |
|                                       | Neural language models, machine translation |
| Multidisciplinary research             | HCI, ML, NLP, programming languages
|                                       | Driving applications |

We need open-world collaborative research!
A Research Course for Beginners

- Hardest part of a PhD: how to select a topic
  - Apprentice under a thesis supervisor
- A true and tried technique for junior researchers
  - Work with a professor, senior graduate students in a small group
  - Choose from an identified research project: meaningful and doable
    - Or suggest a new topic
- Groups of 2 or 3
Course Design

• Background
  • Lectures on basic technology and hands-on experience (2 homeworks)
• Project proposal (Discussions)
  • Proposed research projects in Google docs (on the website)
  • Your ideas are welcome
• 5-week projects
  • Due Mondays: Weekly status updates
  • Tuesday class: small group feedback
  • Thursday class: students take turns in giving mini-lectures on their research topic (an important part of research training)
• Final project presentation and report
# A Tentative Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Tuesday</th>
<th>Thursday</th>
<th>Due (10:30am)</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 7, 9</td>
<td>Course Introduction</td>
<td>Schema → Q&amp;A (HW1)</td>
<td>4/9: Student profile</td>
</tr>
<tr>
<td>April 14, 16</td>
<td>Schema → Dialogues</td>
<td>Tutorial &amp; Discussion (HW2)</td>
<td>4/16: Homework 1</td>
</tr>
<tr>
<td>April 21, 23</td>
<td>Multimodal Assistants</td>
<td>Project Discussions</td>
<td>4/23: Homework 2</td>
</tr>
<tr>
<td>April 28, 30</td>
<td>Project Discussions</td>
<td>ML for NLP Primer</td>
<td>4/30: Project Proposal</td>
</tr>
<tr>
<td>May 5, 7</td>
<td>Group Weekly Meetings</td>
<td>Students’ Mini-lectures</td>
<td></td>
</tr>
<tr>
<td>May 12, 14</td>
<td>Group Weekly Meetings</td>
<td>Students’ Mini-lectures</td>
<td>5/11: Weekly Update</td>
</tr>
<tr>
<td>May 19, 21</td>
<td>Group Weekly Meetings</td>
<td>Students’ Mini-lectures</td>
<td>5/18: Weekly Update</td>
</tr>
<tr>
<td>May 26, 28</td>
<td>Group Weekly Meetings</td>
<td>Students’ Mini-lectures</td>
<td>5/25: Weekly Update</td>
</tr>
<tr>
<td>June 2, 4</td>
<td>Group Weekly Meetings</td>
<td>Students’ Mini-lectures</td>
<td>6/1: Weekly Update</td>
</tr>
<tr>
<td>June 9</td>
<td>Final Project Presentation</td>
<td>—</td>
<td>6/10: Project Report</td>
</tr>
</tbody>
</table>
Grading

- Attendance is mandatory
  - please let us know if you can’t make it to class
- In-class participation: 15%
- Homework: 15%
- Final project: 70%
Let’s Get to Know Each Other
Overview
Conventional Wisdom

• Natural language processing needs a neural network
• Neural network needs well-annotated real users' training data
  • Pre-requisite: Millions of real users
  • Cost: 10,000 Alexa employees annotating real user data
  • Coverage: Millions still don’t have enough coverage
  • Robustness: Dialogue trees, how to handle change of topics?
  • Accuracy: Annotation errors: 30% errors (Multi-Oz)
  • Bootstrapping: How do you start?
  • Scalability: 1.8 B web pages, exponential number of dialogues, thousands of natural language

Metrics: CCRABS
Problem 1

- Will the linguistic technology, web be owned by a duopoly?
  - Alexa: 70% of the 76M installed base of owners in the US
    - 100,000 3rd-party skills, 60,000 compatible IoT devices
- Will it cover the entire web (incl. non-profit)? Rare languages? Is it feasible? Is it profitable?
- Monopolies hurt consumers
  - Privacy, open competition, innovation, quality of service
Protect Privacy with an Open Federated Architecture

• **NLP**
  - training in the cloud (currently)
  - inference locally (in the future)

• **Almond: Privacy-preserving assistant**
  - Keeps users accounts & data local
  - Communicate/share with each other (like email)
  - Users share in natural language

• **Integrated with Home Assistant**

A fully-functional research prototype is available as Almond for Android/web.
Problem 2

- Purely neural approach is prohibitively expensive
Vision of the Future Virtual Assistants

- The entire Web is going voice-accessible!
- Redefine Search
  Based on history, emails, calendar, articulated user preference
- Automation: Natural language programming
  - Personal: order groceries, food every week or evening, pay bills..
  - Doctors, stock brokers, loan officers
- Advisors Behavior influence/manipulation
  - Fitness, bodybuilding, finances, education, careers

We need a new methodology that is open to all!
Search for an upscale restaurant and then make a reservation for it
Reserve a high-end restaurant for me
Can you reserve a restaurant for me? I want an upscale place.

我想预约一个高档餐厅
找一家高档餐厅，然后帮我预约
بگذارید ملاقات قرار من برای وکنید پیدا خوب رستوران یک
Alexa’s 2-Step Approach

Step 1
- Natural Language Commands
- Neural Network
- Alexa Meaning Representation Language (AMRL)

Step 2
- Alexa Meaning Representation Language (AMRL)
- Interpreter
- Execute
Idea 1: End-to-End Translation

- Human-computer communication
  - Easier than understanding human-human communication.
- ThingTalk: formal virtual assistant programming language
  - Capture full capability
  - Independent of language syntax, source natural language
- End-to-end translation
  - Let neural network figure out the intermediate representation

Text

Search for an upscale restaurant and make a reservation for it

Meaning: ThingTalk code

now => @com.yelp.Restaurant()
price == enum(expensive)
=> @com.yelp.reserve(restaurant=id)
Could you please get me a restaurant that is upscale? I want to reserve one.

Reserve me a luxury restaurant

给我们找一家高级餐厅并预约

Eʻimi i kahi hale ʻaina hulahula a laila
hana iā ia no ka mālama ʻana iā ia

高級レストランを検索してから予約する

Che cosa stai cercando? Ho bisogno di qualcosa di lussoso.

Cerca un ristorante di lusso e dammi la prenotazione

Prenotami un ristorante da lusso
Idea 2: Training-Data Engineering

• Tools to address CCRABS
  • cost, coverage, robustness, accuracy, bootstrapping, scalability
• Apply CS engineering approach to AI training data
Q&A

Alexa
User hand-codes question/code 1 by 1

get me an upscale restaurant
What are the restaurants around here?
What is the best restaurant?
search for Chinese restaurants

Genie: Synthesizes question/code from a schema

User

Schema

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
</tr>
</thead>
</table>

Field Annotations

500 Domain-Independent Templates

What is the <prop> of <subject>?
What is the <subject>'s <prop>?

get me an upscale restaurant
What are the restaurants around here?
What is the best restaurant?
search for Chinese restaurants

find restaurants that serve Chinese or Japanese food
What is the best non-Chinese restaurant near here?
Show me a cheap restaurant with 5-star review.
Are there any restaurant with at least 4.5 stars?
What is the phone number of Wendy's?
I'm looking for an Italian fine dining restaurant.
Give me the best Italian restaurant.
Find me the best restaurant with 500 or more reviews
Show me some restaurant with less than 10 reviews

...
Today’s Dialogue Trees: Laborious & Brittle

A: Hello, how can I help you?

U: I'm looking to book a restaurant for Valentine's Day

A: What kind of restaurant?

U: Terun on California Ave
-- or --
U: Something that has pizza
-- or --
U: I don't know, what do you recommend?

NLU: intent + slots

ReserveAction

Domain-specific rule-based policy

ElicitSlot

ShowResults

Recommend

Hard-coded sentences

Name = “Terun”

Food = “pizza”

Fixed set of follow-up intents

???
Alexa: Annotate 1 Dialogue at a Time

Annotation of intents and slots

30% error!
Genie: Transaction Dialogue State Machine

- Init
- Greet
- SearchRequest
  - Greet
  - SlotFillQuestion
  - ProposeRefine
  - ProposeOne
  - ProposeN
- InfoRequest
- ProvideInfo
  - SearchRefine
  - SearchQuestion
  - AskAction
  - InfoQuestion
  - ProvideInfo
  - SlotFillQuestion
  - Answer
    - ConfirmAction
    - ExecuteAction
    - Confirm
    - ActionQuestion
    - Thanks
    - ProvideInfo
    - End
## Technology Stack

### Businesses

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Cuisine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terun</td>
<td>moderate</td>
<td>pizza</td>
</tr>
<tr>
<td>Coconuts</td>
<td>moderate</td>
<td>caribbean</td>
</tr>
</tbody>
</table>

### Sentence Templates

- What is the `<prop>` of `<subject>`?
- What is the `<subject>’s <prop>`?

### Dialogue Models

Transaction Dialogue State Model

### Training Data

<table>
<thead>
<tr>
<th>StateResult</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant, price == moderate &amp; geo == “Palo Alto”</td>
</tr>
<tr>
<td>{ id = “Terun”, price = moderate, cuisines = [“pizza”], … }</td>
</tr>
<tr>
<td>{ id = “Coconuts”, price = moderate, cuisines = [“caribbean”] }</td>
</tr>
</tbody>
</table>

### Neural Network

Restaurant Reservation Agent

---

**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?

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

**ProposeN**: I found Terun and Coconuts. Both are moderately priced.
Contextual Language Understanding Model

Search: @Yelp.Restaurant, price == cheap & & ...
Preliminary Results

**API annotations → multi-domain event-based actions**
When Apple’s stock drops to $200, buy $10,000

**API annotations → Access control**
My dad can view my security camera if I am not home.

**Schema annotations → accurate complex queries**
Find a Spanish restaurant open at 10pm

**Schema annotations → Neural dialogue acts + agent**
61% turn-by-turn accuracy on restaurants in MultiWoz

**Transfer learning to new domains (MultiWoz dialogues)**
Synthesized data training achieves 73% of real data.
# Potential Projects

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applications</td>
<td>Assistants: Social, Music, COVID-19, Minecraft for Autistic Children</td>
</tr>
<tr>
<td>Multi-disciplinary</td>
<td>Two-Way Conversations</td>
</tr>
<tr>
<td>HCI + NLP</td>
<td>Program by Example + Voice</td>
</tr>
<tr>
<td>ML</td>
<td>Improvement with User Feedback</td>
</tr>
<tr>
<td></td>
<td>Neural Model Experimentation for Assistants</td>
</tr>
<tr>
<td></td>
<td>Multi-Lingual Assistants</td>
</tr>
<tr>
<td></td>
<td>Controllable and Natural Response Generation</td>
</tr>
<tr>
<td></td>
<td>Multi-Domain Transactional Dialogues</td>
</tr>
<tr>
<td>Systems</td>
<td>Automatic Template Creation</td>
</tr>
<tr>
<td></td>
<td>Completeness of Template-Based Question Synthesis</td>
</tr>
</tbody>
</table>