Lecture 9: Human conversation, frame-based dialogue systems
Outline

- Basic Conversational Agents
  - ASR
  - NLU
  - Generation
  - Dialogue Manager
- Dialogue Manager Design
  - Finite State
  - Frame-based
  - Information State
    - Dialogue acts and grounding in humans
    - Dialog act generation: confirmation and reject
    - Dialog act interpretation
Conversational Agents

**AKA:**
- Spoken Language Systems
- Dialogue Systems
- Speech Dialogue Systems

**Applications:**
- Travel arrangements (Amtrak, United airlines)
- Telephone call routing
- Tutoring
- Communicating with robots
- Anything with limited screen/keyboard
A travel dialog: Communicator
Xu and Rudnicky (2000)

S1: Hello. You’ve reached the [Communicator]. Tell me your full name
U2: hi i’d like to fly to seattle tuesday morning
S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4: [John Doe]
S5: Hello, [John], I’m glad to hear from you again. You can now ask me for help.
S6: OK. Let’s see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?
U7: Yes
S8: Will you return to Pittsburgh from Seattle?
U9: no i will go to Los Angeles
S10: Traveling to Los Angeles. on what date
U11: Sorry can we go back to talk about the previous leg again
S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?
U13: actually i want to leave on wednesday
S14: a flight on wednesday.
S15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?
U16: Yes
Call routing: ATT HMIHY
Goren et al. (1997)

<table>
<thead>
<tr>
<th>S:</th>
<th>How may I help you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>U:</td>
<td><em>can you tell me how much it is to tokyo?</em></td>
</tr>
<tr>
<td>S:</td>
<td>You want to know the cost of a call?</td>
</tr>
<tr>
<td>U:</td>
<td><em>yes that’s right</em></td>
</tr>
<tr>
<td>S:</td>
<td>Please hold on for rate information.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S:</th>
<th>How may I help you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>U:</td>
<td><em>what’s the area code for chicago?</em></td>
</tr>
<tr>
<td>S:</td>
<td>You want an area code.</td>
</tr>
<tr>
<td>U:</td>
<td><em>(silence)</em></td>
</tr>
<tr>
<td>S:</td>
<td>The area code for Chicago is 312.</td>
</tr>
</tbody>
</table>
A tutorial dialogue: ITSPVOKE
Litman and Silliman (2004)

S: I didn’t understand. What force acts on all objects within close proximity to earth?

U: *uh the force of gravity*

S: Fine. Besides the Earth’s gravitational force, are there any other forces acting on the person, keys, and elevator after the release?

U: *no*

S: Yep. Thus, keep in mind that after the release, the only force on the person, keys, and elevator is the force of gravity. If gravity is the only force acting on an object, will it be moving or staying still?

U: *it will be moving*

S: Fine. Not only are the person, keys, and elevator moving, they have only gravitational forces on them. When an object is falling and has only gravitational force on it, it is said to be in what?

U: *freefall*
“Make an appointment for Tuesday”

What time is your appointment?

“215”

OK, I can create your meeting. Note that you already have an appointment at 2:15 pm. Shall I schedule it anyway?

“No”

To continue, you can Confirm, Cancel, Change the Time, or Change the Title.

OK. You’re probably way too busy anyway, Dan.
"Find restaurants near me"

I found fifteen restaurants fairly close to you:

**15 Restaurants**

Emmy's Spaghetti Shack 0.2 mi
18 Virginia Ave
Italian, $$$$$
★★★★★ 1101 Reviews

ICHI Sushi 0.2 mi
3369 Mission St
Japanese, Sushi Bars, $$$$$
★★★★★ 260 Reviews

Avedano's Holly Park M... 0.2 mi

"Tell me more about the second one"

I'm sorry, Dan, I'm afraid I can't do that.
“Find restaurants near me”
  tap to edit

I found fifteen restaurants fairly close to you:

15 Restaurants

**Emmy's Spaghetti Shack**  0.2 mi
18 Virginia Ave
Italian, $$$$$
★★★★★ 1101 Reviews

**ICHI Sushi**  0.2 mi
3369 Mission St
Japanese, Sushi Bars, $$$$$
★★★★★ 260 Reviews

**Avedano's Holly Park M...**  0.2 mi

“Are any of them Italian”
  tap to edit

My web search turned this up:

**Web Search**

Are any of them Italian

any - Dizionario inglese-italiano
WordReference
www.wordreference.com

English-Italian Dictionary | any ... of any sort
adj (of an unspecified variety) di qualsiasi

Italian language - Wikipedia, the free encyclopedia
en.wikipedia.org
Italian or lingua italiana) is a Romance
Personal Assistants

- Siri
- Google Now
- Microsoft Cortana, etc
Dialogue System Architecture

- Speech Recognition
- Natural Language Understanding
- Text-to-Speech Synthesis
- Natural Language Generation
- Dialogue Manager
- Task Manager
Dialog architecture for Personal Assistants

Bellegarda
Dialogue Manager

- Controls the architecture and structure of dialogue
- Takes input from ASR/NLU components
- Maintains some sort of state
- Interfaces with Task Manager
- Passes output to NLG/TTS modules
Four architectures for dialog management

Finite State
Frame-based
Information State (Markov Decision Process)
Classic AI Planning
Finite-State Dialog Management

Consider a trivial airline travel system:
  
  Ask the user for a departure city
  
  Ask for a destination city
  
  Ask for a time
  
  Ask whether the trip is round-trip or not
Finite State Dialog Manager

1. What city are you leaving from?
2. Where are you going?
3. What date do you want to leave?
4. Is it a one-way trip?
   - Yes: Do you want to go from <FROM> to <TO> on <DATE>?
   - No: What date do you want to return?
     - Yes: Do you want to go from <FROM> to <TO> on <DATE> returning on <RETURN>?
     - No: Book the flight
Finite-state dialog managers

- System completely controls the conversation with the user.
- It asks the user a series of questions
- Ignoring (or misinterpreting) anything the user says that is not a direct answer to the system’s questions
• Systems that control conversation like this are **system initiative** or **single initiative**.

• **Initiative**: who has control of conversation

• In normal human-human dialogue, initiative shifts back and forth between participants.
System Initiative

System completely controls the conversation

- Simple to build
- User always knows what they can say next
- System always knows what user can say next
  - Known words: Better performance from ASR
  - Known topic: Better performance from NLU
- OK for VERY simple tasks (entering a credit card, or login name and password)

- Too limited
Problems with System Initiative

- Real dialogue involves give and take!
- In travel planning, users might want to say something that is not the direct answer to the question.
- For example answering more than one question in a sentence:
  Hi, I’d like to fly from Seattle Tuesday morning I want a flight from Milwaukee to Orlando one way leaving after 5 p.m. on Wednesday.
Single initiative + universals

- We can give users a little more flexibility by adding **universals**: commands you can say anywhere
- As if we augmented every state of FSA with these
  
  Help
  
  Start over
  
  Correct

- This describes many implemented systems
- But still doesn’t allow user much flexibility
User Initiative

- User directs the system
  - Asks a single question, system answers
- Examples: **Voice web search**
- But system can’t:
  - ask questions back,
  - engage in clarification dialogue,
  - engage in confirmation dialogue
Mixed Initiative

• Conversational initiative can shift between system and user

• Simplest kind of mixed initiative: use the structure of the **frame** to guide dialogue
An example of a frame

FLIGHT FRAME:

ORIGIN:

  CITY: Boston
  DATE: Tuesday
  TIME: morning

DEST:

  CITY: San Francisco

AIRLINE:

  ...

Mixed Initiative

- Conversational initiative can shift between system and user
- Simplest kind of mixed initiative: use the structure of the frame to guide dialogue

<table>
<thead>
<tr>
<th>Slot</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORIGIN</td>
<td>What city are you leaving from?</td>
</tr>
<tr>
<td>DEST</td>
<td>Where are you going?</td>
</tr>
<tr>
<td>DEPT DATE</td>
<td>What day would you like to leave?</td>
</tr>
<tr>
<td>DEPT TIME</td>
<td>What time would you like to leave?</td>
</tr>
<tr>
<td>AIRLINE</td>
<td>What is your preferred airline?</td>
</tr>
</tbody>
</table>
Frames are mixed-initiative

- User can answer multiple questions at once.
- System asks questions of user, filling any slots that user specifies
  - When frame is filled, do database query
- If user answers 3 questions at once, system has to fill slots and not ask these questions again!
  - Avoids strict constraints on order of the finite-state architecture.
Multiple frames

- flights, hotels, rental cars
- Flight legs: Each flight can have multiple legs, which might need to be discussed separately
- Presenting the flights (If there are multiple flights meeting users constraints)
  - It has slots like 1ST_FLIGHT or 2ND_FLIGHT so user can ask “how much is the second one”
- General route information:
  - Which airlines fly from Boston to San Francisco
- Airfare practices:
  - Do I have to stay over Saturday to get a decent airfare?
Natural Language Understanding

- There are many ways to represent the meaning of sentences
- For speech dialogue systems, most common is “Frame and slot semantics”.
An example of a frame

Show me morning flights from Boston to SF on Tuesday.

SHOW:

FLIGHTS:

ORIGIN:
  CITY: Boston
  DATE: Tuesday
  TIME: morning

DEST:
  CITY: San Francisco
Semantics for a sentence

LIST  FLIGHTS  ORIGIN
Show me flights from Boston

DESTINATION  DEPARTDATE
to San Francisco on Tuesday

DEPARTTIME
morning
Idea: HMMs for semantics

- Hidden units are slot names
  - ORIGIN
  - DESTCITY
  - DEPARTTIME

- Observations are word sequences on Tuesday
HMM model of semantics

Pieraccini et al (1991)
Semantic HMM

• Goal of HMM model:
  To compute labeling of semantic roles $C = c_1, c_2, ..., c_n$
  (C for ‘cases’ or ‘concepts’)
  that is most probable given words $W$

$$\arg\max_C P(C \mid W) = \arg\max_C \frac{P(W \mid C)P(C)}{P(W)}$$

$$= \arg\max_C P(W \mid C)P(C)$$

$$= \arg\max \prod_{i=2}^{N} P(w_i \mid w_{i-1} \ldots w_1, C)P(w_1 \mid C) \prod_{i=2}^{M} P(c_i \mid c_{i-1} \ldots c_1)$$
Semantic HMM

• From previous slide:

\[
= \underset{c}{\text{argmax}} \prod_{i=2}^{N} P(w_i \mid w_{i-1}...w_1,C) P(w_1 \mid C) \prod_{i=2}^{M} P(c_i \mid c_{i-1}...c_1)
\]

• Assume simplification:

\[
P(w_i \mid w_{i-1}...w_1,C) = P(w_i \mid w_{i-1},...,w_{i-N+1},c_i)
\]

\[
P(c_i \mid c_{i-1}...c_1,C) = P(c_i \mid c_{i-1},...,c_{i-M+1})
\]

• Final form:

\[
= \underset{c}{\text{argmax}} \prod_{i=2}^{N} P(w_i \mid w_{i-1}...w_{i-N+1},c_i) \prod_{i=2}^{M} P(c_i \mid c_{i-1}...c_{i-M+1})
\]
semi-HMM model of semantics

Pieraccini et al (1991)

\[
P(W|C) = P(\text{me}|\text{show,SHOW}) P(\text{show}|\text{SHOW}) P(\text{flights}|\text{FLIGHTS})... P(\text{FLIGHTS}|\text{SHOW}) P(\text{DUMMY}|\text{FLIGHTS})...
\]
Semi-HMMs

- Each hidden state
  - Can generate multiple observations
- By contrast, a traditional HMM
  - One observation per hidden state
  - Need to loop to have multiple observations with the same state label
How to train

- Supervised training
- Label and segment each sentence with frame fillers
- Essentially learning an N-gram grammar for each slot

```
LIST      FLIGHTS   DUMMY ORIGIN   DEST
Show me   flights    that go    from Boston to SF
```
Another way to do NLU: Semantic Grammars

- CFG in which the LHS of rules is a semantic category:

  LIST -> show me | I want | can I see | ...
  DEPARTTIME -> (after | around | before) HOUR
    | morning | afternoon | evening
  HOUR -> one | two | three... | twelve (am | pm)
  FLIGHTS -> (a) flight | flights
  ORIGIN -> from CITY
  DESTINATION -> to CITY
  CITY -> Boston | San Francisco | Denver | Washington
Tina parse tree with semantic rules

Seneff 1992
Phoenix SLU system: Recursive Transition Network

Ward 1991, figure from Wang, Deng, Acero
A final way to do NLU: Condition-Action Rules

- Active Ontology: relational network of concepts
- **data structures**: a *meeting* has
  - a date and time,
  - a location,
  - a topic
  - a list of attendees
- **rule sets** that perform actions for concepts
  - the *date* concept turns string
    - *Monday at 2pm* into
    - *date object* `date(DAY,MONTH,YEAR,HOURS,MINUTES)`
Rule sets

- Collections of **rules** consisting of:
  - condition
  - action

- When user input is processed, facts added to store and
  - rule conditions are evaluated
  - relevant actions executed
Part of ontology for meeting task

meeting concept: if you don’t yet have a location, ask for a location
Other components
ASR: Language Models for dialogue

- Often based on hand-written Context-Free or finite-state grammars rather than N-grams

- Why?
  - Need for understanding; we need to constrain user to say things that we know what to do with.
ASR: Language Models for Dialogue

- We can have LM specific to a dialogue state
- If system just asked “What city are you departing from?”
- LM can be
  - City names only
  - FSA: (I want to (leave | depart)) (from) [CITYNAME]
  - N-grams trained on answers to “Cityname” questions from labeled data
- A LM that is constrained in this way is technically called a “restricted grammar” or “restricted LM”
Generation Component

- **Content Planner**
  - Decides what content to express to user
    (ask a question, present an answer, etc)
  - Often merged with dialogue manager

- **Language Generation**
  - Chooses syntax and words

- **TTS**

- **In practice**: Template-based w/most words prespecified
  
  What time do you want to leave CITY-ORIG?
  Will you return to CITY-ORIG from CITY-DEST?
More sophisticated language generation component

- Natural Language Generation
- Approach:
  - Dialogue manager builds representation of meaning of utterance to be expressed
  - Passes this to a “generator”
- Generators have three components
  - Sentence planner
  - Surface realizer
  - Prosody assigner
Architecture of a generator for a dialogue system

Walker and Rambow 2002)
HCI constraints on generation for dialogue: “Coherence”

Discourse markers and pronouns (“Coherence”):

Please say the date.

Please say the start time.

Please say the duration...

Please say the subject...

First, tell me the date.

Next, I’ll need the time it starts.

Thanks. <pause> Now, how long is it supposed to last?

Last of all, I just need a brief description
HCI constraints on generation for dialogue: coherence (II): tapered prompts

Prompts which get incrementally shorter:

System: Now, what’s the first company to add to your watch list?
Caller: Cisco

System: What’s the next company name? (Or, you can say, “Finished”)
Caller: IBM

System: Tell me the next company name, or say, “Finished.”
Caller: Intel

System: Next one?
Caller: America Online.
System: Next?
Caller: ...
How mixed initiative is usually defined

- First we need to define two other factors
  - Open prompts vs. directive prompts
  - Restrictive versus non-restrictive grammar
Open vs. Directive Prompts

- **Open prompt**
  - System gives user very few constraints
  - User can respond how they please:
    “How may I help you?” “How may I direct your call?”

- **Directive prompt**
  - Explicit instructs user how to respond
    “Say yes if you accept the call; otherwise, say no”
Restrictive vs. Non-restrictive grammars

- Restrictive grammar
  - Language model which strongly constrains the ASR system, based on dialogue state

- Non-restrictive grammar
  - Open language model which is not restricted to a particular dialogue state
Definition of Mixed Initiative

<table>
<thead>
<tr>
<th>Grammar</th>
<th>Open Prompt</th>
<th>Directive Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrictive</td>
<td><em>Doesn’t make sense</em></td>
<td><em>System Initiative</em></td>
</tr>
<tr>
<td>Non-restrictive</td>
<td><em>User Initiative</em></td>
<td><em>Mixed Initiative</em></td>
</tr>
</tbody>
</table>
Evaluation

1. Slot Error Rate for a Sentence
   \[ \frac{\text{# of inserted/deleted/substituted slots}}{\text{# of total reference slots for sentence}} \]

2. End-to-end evaluation (Task Success)
Evaluation Metrics

“Make an appointment with Chris at 10:30 in Gates 104”

<table>
<thead>
<tr>
<th>Slot</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>Chris</td>
</tr>
<tr>
<td>TIME</td>
<td>11:30 a.m.</td>
</tr>
<tr>
<td>ROOM</td>
<td>Gates 104</td>
</tr>
</tbody>
</table>

Slot error rate: 1/3
Task success: At end, was the correct meeting added to the calendar?
Linguistics of Human Conversation

- Turn-taking
- Speech Acts
- Grounding
Turn-taking

Dialogue is characterized by turn-taking.

A:
B:
A:
B:
...

So how do speakers know when to take the floor?
Adjacency pairs

Sacks et al. (1974)

- **Adjacency pairs**: current speaker selects next speaker
  - Question/answer
  - Greeting/greeting
  - Compliment/downplayer
  - Request/grant

- Silence inside the pair is meaningful:

  A: Is there something bothering you or not?  
  *(1.0)*  
  A: Yes or no?  
  *(1.5)*  
  A: Eh  
  B: No.
Speech Acts

- Austin (1962): An utterance is a kind of action
- Clear case: performatives
  I name this ship the Titanic
  I second that motion
  I bet you five dollars it will snow tomorrow
- Performative verbs (name, second)
- Austin’s idea: not just these verbs
5 classes of “speech acts”

**Assertives:** committing the speaker to something’s being the case (suggesting, putting forward, swearing, boasting, concluding)

**Directives:** attempts by speaker to get addressee to do something (asking, ordering, requesting, inviting, advising, begging)

**Commissives:** Committing speaker to future course of action (promising, planning, vowing, betting, opposing)

**Expressives:** expressing psychological state of the speaker about a state of affairs (thanking, apologizing, welcoming, deploring).

**Declarations:** changing the world via the utterance (I resign; You’re fired)

Searle (1975)
More Illocutionary acts: Grounding

• Why do elevator buttons light up?

• Clark (1996) (after Norman 1988)
  
  **Principle of closure.** Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it

• What is the linguistic correlate of this?
Grounding

• Need to know whether an action succeeded or failed
• Dialogue is also an action
  • a collective action performed by speaker and hearer
  • Common ground: set of things mutually believed by both speaker and hearer
• Need to achieve common ground, so hearer must ground or acknowledge speakers utterance.
How do speakers ground?
Clark and Schaefer

- **Continued attention:**
  - B continues attending to A

- **Relevant next contribution:**
  - B starts in on next relevant contribution

- **Acknowledgement:**
  - B nods or says continuer (*uh-huh*) or assessment (*great!*)

- **Demonstration:**
  - B demonstrates understanding A by **reformulating** A’s contribution, or by **collaboratively completing** A’s utterance

- **Display:**
  - B repeats verbatim all or part of A’s presentation
A human-human conversation

C1: ...I need to travel in May.
A1: And, what day in May did you want to travel?
C2: OK uh I need to be there for a meeting that’s from the 12th to the 15th.
A2: And you’re flying into what city?
C3: Seattle.
A3: And what time would you like to leave Pittsburgh?
C4: Uh hmm I don’t think there’s many options for non-stop.
A4: Right. There’s three non-stops today.
C5: What are they?
A5: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C6: OK I’ll take the 5ish flight on the night before on the 11th.
C7: OK.
Grounding examples

Display:

C: I need to travel in May
A: And, what day in May did you want to travel?

Acknowledgement

C: He wants to fly from Boston
A: mm-hmm
C: to Baltimore Washington International
Grounding Examples (2)

- Acknowledgement + next relevant contribution
  And, what day in May did you want to travel? And you’re flying into what city? And what time would you like to leave?
- The and indicates to the client that agent has successfully understood answer to the last question.
Grounding negative responses
From Cohen et al. (2004)

• System: Did you want to review some more of your personal profile?
• Caller: No.
• System: Okay, what’s next?

• System: Did you want to review some more of your personal profile?
• Caller: No.
• System: What’s next?
Summary

- The Linguistics of Conversation
- Basic Conversational Agents
  - ASR
  - NLU
  - Generation
  - Dialogue Manager
- Dialogue Manager Design
  - Finite State
  - Frame-based
  - Initiative: User, System, Mixed