Advanced Dialog Workshop

What’s “Advanced Dialog”
Marie Meteer
Brandeis University

March 3rd, 2014
Outline

- Workshop Work
  - Features of the “conference personal assistant”

- What makes a dialog “advanced”
  - Walk through an example

- Architectures
  - Where we are and where we need to go

- What can linguistics tell us

- No interpretation without representation!
  - Making sure our back end aligns with discourse goals

- NLU: From words to the representation
  - Back to intention and referring expressions
Conference Personal Assistant

- Ready access to all conference schedule information
  - By time, speaker, speaker’s company, track, location
  - Find out about times and locations for breaks, keynotes, lunch, and the reception

- Ability to interact with a personal schedule with the sessions you want to attend
  - Adding directly from the conference schedule to your personal schedule by voice or touch
  - Querying for particular speakers and adding sessions where they are speaking
  - Querying where and when sessions you scheduled will be
What makes an “Advanced” Dialog System

- **Mixed initiative**
  - User can change the topic or revisit previous dialog elements
  - System can take control to get more clarifying information
  - User can say anything at any point and get some intelligent response

- **Multimodal**
  - Input and output can be audio/visual/tactile
  - Linked gesture and speech (e.g. “When is that one” or “I want to go to the third one”)

- **“Natural” interaction, variability of expression**
  - “Persona”: Chatty and friendly, businesslike

- **“Self aware”**
  - System identifies the confidence of an interpretation
  - Knows when to ask for clarification and when to move forward
  - Knows when it's being asked something beyond the applications capability

- **Helpful**
  - Recognizes users intentions
  - Offers additional information or incorporates multiple steps
Critical Element: Context

- Context aware
  - “Remembers” what the user has already said and uses that information
  - Tracks the users “focus” so speakers can refer back to objects already mentioned
  - Recognizes the users goals based on what has happened so far
  - Recognizes visual and gesture as contextual elements
What does context give us?

- Narrow the search
  - Show me the restaurants near here.
  - Which ones are open?

- Pointing with words
  - The second one
  - The Indian one

- Tracking objects mentioned already
  - Reuse for new tasks
  - Going back to an unfinished task

- Track user goals to understand tasks in context

- Ability to “recover” when user is misunderstood or user changes his/her mind
  - Guide “clarification” dialogs
Outline

- Workshop Work
  - Features of the “conference personal assistant”

- What makes a dialog advanced
  - Walk through an example

- Architectures
  - Where we are and where we need to go

- What can linguistics tell us

- No interpretation without representation!
  - Making sure our back end aligns with discourse goals

- NLU: From words to the representation
  - Back to intention and referring expressions
Traditional Speech IVR

- Dictionary
- Acoustic model
- Grammar
- Call Flow
- Actions / Backend

- Finite state grammar defines all word sequences and meanings
- Prerecorded speech or TTS

Context:
- Defined by call flow
- Next action Defined by call flow

© MM Consulting 2014
“NextGen” added “NLU”

- Dictionary
- Acoustic model
- SLM, Topic ID, Partial Parsing
- Call Flow
- Task Manager
- Actions / Backend

- SR with SLM
- SSM: one label per model
- NLU: Grammar fills slot values
- Prerecorded speech or TTS
- Next action
- Defined by call flow

- Context:
  - Defined by call flow

© MM Consulting 2014
Voice Search

SR with SLM → Search term extraction → Form query

- No Context (except maybe ordered autocomplete)
- Format results

- Prerecorded speech or TTS, Text to screen

Dictionary

SLM

Info Extraction

Actions / Backend

Acoustic model

Voice Data
Clicks

Voice Text
Images
Highlighting
Dialogue System Architecture
(ala MIT Galaxy, 1994)

- Dictionary
- Acoustic model
- Grammar/SLM
- Parser/Topic ID/NLU
- Dialog strategies Task models
- Actions / Backend
- Speech Recognition
- Natural Language Understanding
- Dialog Manager
  - Context update
  - Response Strategy Selection
- Speech Synthesis (TTS)
- Natural Language Generation
- Task Manager

Voice Data
Clicks

Text Images
Highlighting

© MM Consulting 2014
The ejTalk Approach
(real diagram to come)

- Acoustic model
- Dictionary
- Grammar/SLM
- Parser/Topic ID/NLU
- Dialog strategies
- Task models
- Actions/Backend

- Speech Recognition SR engines
- NLU Engines
- Context update
- Conversation Manager
- Task Manager

- Natural Language Generation
- Response Strategy Selection

Voice Data
Clicks

text

Voice Text
Images
Highlighting

© MM Consulting 2014
Examples of where we want to go

U: When am I speaking?
S: 2 pm on Tuesday
U: Put that on my schedule
U: Is anyone from Sensory speaking?
S: Yes. Jeff Rogers and Todd Mozer.
U: What is Todd speaking on?
S: He’s on the panel. Here’s the description.
S: Would you like that on your personal schedule
U: Thanks. How about Jeff?
Continued Examples

U: How about Jeff?

S: Jeff Rogers is speaking on Truly Handsfree at 11 am on Tuesday.

U: (notices time) Do I have a session scheduled now?

S: No, until 11 am.

U: OK. What’s in the other track?

S: Here are the talks in the Business track at 11: <shows results>

U: I’d like to hear <Pat>. Schedule me for that one.

S: ??

• Ask to repeat
• Ask if it’s Matt?
• Ignore misreco and execute command
What’s in the “Discourse Manager” box?

- Context created by the discourse so far
  - Allows “abbreviated” sentences
    - How much do I owe [on my visa bill]

- Expectation of what might happen next
  - Task representation with sequences of events (Agenda)
  - Proactively offer of next steps

- Knowledge of the tasks the system is capable of
  - Connection of tasks to backend systems
  - Representation of what can’t be done

- Knowledge of the input and output modalities and how they relate
  - Need to know what’s in the list to act on “next one”
What’s new, What’s not

- New
  - Zillions of apps
  - Rarely more than inches from significant computing power in our phones and laptops
  - Nearly ubiquitous connectivity to “the cloud”

- Not new
  - Theory of discourse, plans, referring expressions

  **We don’t have to figure this out from scratch!!!**
Outline

- Workshop Work
  - Features of the “conference personal assistant”

- What makes a dialog advanced
  - Walk through an example

- Architectures
  - Where we are and where we need to go

- What can linguistics tell us

- No interpretation without representation!
  - Making sure our back end aligns with discourse goals

- NLU: From words to the representation
  - Back to intention and referring expressions
Linguistics of Human Conversation

- Turn-taking
  - Who gets to talk next

- Grounding
  - Feedback on where in the conversation we are

- Implicature
  - Understanding more than is said
  - “Cooperative principle

- Conversational Structure
Grounding

- Why do elevator buttons light up?

- 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?
  - Repeating all or part of an utterance:
    - C: I need to travel in May
    - A: And, what day in May did you want to travel?
  - Backchannel “uh-huh”, “right”
  - Closure “OK, Now ...."
Conversational Implicature

Agent: And, what day in May did you want to travel?

Client: My meeting goes from the 12th to the 15th.

- Note that client did not answer question.
- What is it that licenses agent to infer that client is mentioning this meeting so as to inform the agent of the travel dates?
- Cooperative Principle (Grice 1975)
  - This is a tacit agreement by speakers and listeners to cooperate in communication
4 Gricean Maxims

- **Relevance**: Be relevant
  - In the last example the Hearer thinks:
    - Speaker is following maxims, would only have mentioned meeting if it was relevant. How could meeting be relevant? Must need to be there in time for the meeting

- **Quantity**: Do not make your contribution more or less informative than required
  - A: Did you do the reading for today’s class?
  - B: I intended to
    - B’s answer would be true if B intended to do the reading AND did the reading, but would then violate maxim

- **Quality**: try to make your contribution one that is true
  - Don’t say things that are false or for which you lack adequate evidence

- **Manner**: Avoid ambiguity and obscurity; be brief and orderly
Discourse Structure
Grosz and Sidner, 1986

- Three components
  - Linguistic: sentence, phrase structure, cue phrases
  - Attentional: Focus of attention
  - Intentional: Speaker’s purpose and the relations between those purposes

- “Structure” is key
  - Originally described as a stack
  - Really needs to be a persistent representation
    - Once said, things never really go away
What is “Intentional” Structure

- **What is the speaker trying to do?**

- Intention, plans, goals

- **Tasks**
  - Ravenclaw
What is “Attentional” Structure

What are they talking about?

- Focus, anaphora
- Centering
  - Walker, Marilyn A. "Centering, anaphora resolution, and discourse structure." *Centering theory in discourse (1998)*
Referring Expressions

• Referring” goes beyond objects
  • According to Doug, Sue just bought a 1962 Ford Falcon
    • But that turned out to be a lie (a speech act)
    • But that was false (proposition)
    • That struck me as a funny way to describe the situation (manner of description)
    • That caused Sue to become rather poor (event)

• And beyond speech
  • Show the restaurants near here.
    • (select one with cursor)
    • Is that one open tonight?
    • Is it open tomorrow?
Outline

- Workshop Work
  - Features of the “conference personal assistant”

- What makes a dialog advanced
  - Walk through an example

- Architectures
  - Where we are and where we need to go

- What can linguistics tell us

- No interpretation without representation!
  - Making sure our back end aligns with discourse goals

- NLU: From words to the representation
  - Back to intention and referring expressions
No interpretation without representation!

• Representation challenges
  • Many ways to refer to a session
    • I can ask for a session by time, speaker, type of session
    • I can refer to a session with respect to another session
    • The next session, the session in the other track, the first session, etc.
  • Some ways of referring are unambiguous, e.g. “2 pm Tuesday, business track”
  • Others are ambiguous, e.g. speakers talking in multiple sessions
  • Names of talks and people can vary
    • “Truly Handsfree today and Tomorrow” vs “Truly Handsfree”
    • William Miesel vs. Bill (though this may be more easily handed with explicit nicknames)
Bootstrapping: 1<sup>st</sup> level
Know your App

Out
But predictable
Define specific repair strategies

In the App
Define Intentional Attentional Structure

HUH?
Define Repair Strategies
Outline

- Workshop Work
  - Features of the “conference personal assistant”

- What makes a dialog advanced
  - Walk through an example

- Architectures
  - Where we are and where we need to go

- What can linguistics tell us

- No interpretation without representation!
  - Making sure our back end aligns with discourse goals

- NLU: From words to the representation
  - Back to intention and referring expressions
NLU

- Speech is done (well, not really)
- Need to understand the relationship between the many ways people say things and the tasks and objects in the application
- Back to intention and attention!
- Creating a mapping in advance
- Capturing the dialog context in real time
Intentions and Tasks

- Users Intentions line up with what tasks the system is capable of
  - Find out information about the conference schedule (CS)
  - Use that information to decide what sessions to put on their personal schedule (PS)
  - Query the PS to find out where they should be

- As the dialog progresses, these tasks (and their steps) create the “intentional” structure

- Objects that are named form the “attentional structure”
Intentional and Attentional Structure

Intentional

When am I speaking?  
Put that on my schedule

Tuesday at 2 pm

Attentional

Q-CS

Get-time

Add to PS

Provide DT

Q-CS

[when speaker]  
[speaker MM]  

[session]  
[DT: 3/5 2pm]

Domain

{SESSION 2491
DT:  
Speakers:...  
track ...}
Where in the Architecture?

- Speech recognition uses a statistical language model
  - User can express their requests many different ways
  - Does not have to fit a specific “grammar”

- Statistical task / intention recognition
  - System recognizes what the user is trying to do
    - Get balance, pay bill, view transactions
  - Intentions have to connect to domain tasks

- Information extraction
  - System picks out the specific pieces of information to accomplish the task
    - May need to have some logic to resolve

© MM Consulting 2014
Work to be done

Who’s speaking in the business track after lunch

Task = QueryCS

Sessions:Track2

DateTime= ResolveRef

- Recognize the words
- Identify the intent
  - Find info on sessions
  - Answer is a “who”
- Identify the entities
  - Restriction1 on the session: Track 2
  - Restriction2: Time concept to be resolve (need
NLU and Advanced Dialog Runtime Requirements

- **Topic model**: set of sentences for each “intention”
- **Entity recognizer**: “semantic parsing” rules
- **Relationship recognizer**: role of each entity in the intention.
- **Conceptual representation**: Intention-> tasks, Entities-> slot values
- **Domain representation**: Tasks, Slots, Value types

**SLM**: From list of sentences in the domain

**UI Design**: What are the visual elements available

**Speech Recognition**

**Touch / Gesture**

**Screen Display**

**Audio / TTS**

**Natural Language Understanding**

**Display, prompt selection, Natural Language Generation**

**Dialog Manager**

- Map intention to tasks
- Update context
- Response Strategy Selection

**Response strategies**, information requests, clarification, error handling

© MM Consulting 2014
Integrated components

- Not a strictly sequential process
- Components constrain and feed each other
How to build it?

- EjTalk! (stay in your seats)

- How do I find people that can do this stuff?
  - Brandeis University Computational Linguistics MA
  - [www.brandeis.edu/programs/comp-linguistics/](http://www.brandeis.edu/programs/comp-linguistics/)
  - Or contact me at mmeteer@mac.com

**Thanks!**