Assessing Dialog Management Systems

Presenters:
K.W. (Bill) Scholz
Marie Meteer

With help from the rest of the AVIOS Advanced Dialog Group:
Emmett Coin
John Tadlock
Lorin Wilde
From ASR to Dialog Management

- ASR – speech to text
- NLU – identify intents and concepts
- Context – beyond the sentence or or utterance
- Virtual assistant / personal assistant / digital assistant
- Language User Interface – voice control of app
- DM – integrate ASR, NLU, Context, VA, LUI, …
Dialog System Architecture

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

- **Speech Recognition**
- **Natural Language Understanding**

- **Speech Synthesis (TTS)**
- **Natural Language Generation**

- **Context Update**
- **Response Strategy Selection**

- **Dialog Manager**

- Voice Data Clicks
- Voice Text Images Highlighting
+ **Integrated components**

- Not a strictly sequential process
- Components constrain and feed each other

Speech, NLP & other modalities

Which entities can be referred to with a pronoun

Vocabulary Verb / argument relationships

Domain/Tasks

ejTalk BB

Discourse Representation

Which slots a spoken entity fills

© MM Consulting 2014
Assessing Dialog Managers

What metrics should we use for evaluation of dialog managers?

What are the components of dialog managers to be evaluated?

What aspects of an application need to be managed?

What does the dialog manager do?
+ Mapping the space

- What are the dimensions of complexity for applications
- What are the requirements each dimension puts on a dialog manager
- Now we can design evaluation metrics
Dimensions of application complexity

- **Task**: How complex is each task, single step, requires multiple steps? How many tasks are there?

- **Multitask**: Are tasks completed sequentially or can there be multiple tasks “live” at the same time?

- **Domain**: Does the app do one thing or cover multiple domains, such as managing many different applications?

- **Modality**: Only speech or type/click or multiple modes for input and output?

- **Connectivity and “range”**: Self contained app on a single device or multiple portals into the internet of things?

- **Persistence over time**: Complete “transactions” in a single session or able to come back and continue on the task?
## Dimensions of application complexity:

### Number and complexity of tasks

<table>
<thead>
<tr>
<th>#</th>
<th>Single task, Single step</th>
<th>Related tasks, Clear set of sequential steps</th>
<th>Sequential steps allowing variability</th>
<th>Multiple tasks with multiple steps with dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn on the lights</td>
<td>Recipe or steps to log in/set up account</td>
<td>Shopping, filling out an form</td>
<td>Book a vacation</td>
<td></td>
</tr>
</tbody>
</table>
Dimensions of application complexity:

Number of Domains

<table>
<thead>
<tr>
<th>Single Domain</th>
<th>Fixed set of built in domains</th>
<th>Can “learn” new domains, but need to specify which domain</th>
<th>Open to new domains, system determines domain from discourse context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music player</td>
<td>Siri (treating web search as a single domain)</td>
<td>Alexa skills</td>
<td>My future kitchen assistant</td>
</tr>
</tbody>
</table>
**Dimensions of application complexity:**

### Goal Driven

<table>
<thead>
<tr>
<th>Only task execution</th>
<th>System offers “typical goal”. Simple sequence to meet goal</th>
<th>System can infer user’s goals and mediate competing goals</th>
<th>Multiple actors with shared goals and negotiated goals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Music player</strong></td>
<td>Waze: “Do you want to go to work?” Goal: be at work as soon possible. Waze determines best path.</td>
<td>Recognizes goal from “find the nearest restaurant” and offers to make a reservation (but doesn’t offer to make a reservation at the nearest gas station)</td>
<td>Recognizes from big data what user might want to do next and offers up possibilities. Multiple stakeholders</td>
</tr>
</tbody>
</table>
## Dimensions of application complexity:

**Multimodality**

<table>
<thead>
<tr>
<th># Domain</th>
<th>Single mode (e.g. speech or typing)</th>
<th>Fixed modes for specific actions</th>
<th>Simultaneous synergistic use of multiple modes</th>
<th>Automatic modality transitions depending on device, context,…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>Music player: You speak, it plays music</td>
<td>Voice web query: You speak, it shows you the results</td>
<td>“I want this one” “Show me the blue top” “What’s that”? (takes a picture)</td>
<td>Moving from my car to my house on my phone and now info shows up on the screen.</td>
</tr>
</tbody>
</table>
### Dimensions of application complexity:

#### Reach into the IoT

<table>
<thead>
<tr>
<th># Devices</th>
<th>Control of devices “preprogrammed”</th>
<th>Disambiguating many different IoT devices</th>
<th>Tracks goals and offers to control devices and makes recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only one “thing” on the internet</td>
<td>“Turn on the living room lights” (device and name preprogrammed)</td>
<td>“Turn on the lights” can figure out what house/room you’re in</td>
<td>“Printing your boarding pass. Would you like me to turn down the heat while your gone?”</td>
</tr>
<tr>
<td>Music player: You speak, it plays music</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
+ Dimensions of Applications

# Tasks
Multitasking
Domains
Goal Driven
Multimodal

# Devices
“Reach” / IoT
Personalization
Mxd Initiative
Self aware

# Participants
Measuring complexity

# Tasks
Multitasking
Domains
Goal Driven
Multimodal

# Devices
“Reach” / IoT
Personalization
Mxd Initiative
Self aware

# Participants
+ The Echo on my kitchen counter

# Tasks
Multitasking
Domains
Goal Driven
Multimodal

# Devices
“Reach” / IoT
Personalization
Mxd Initiative
Self aware

# Participants
My Dream Echo with “Future Alexa”

# Tasks
Multitasking
Domains
Goal Driven
Multimodal

# Devices
“Reach” / IoT
Personalization
Mxd Initiative
Self aware

# Participants
Mapping from App capabilities to DM requirements

- As Applications become more complex along these dimensions, the dialog manager needs more capabilities

- Multi-step tasks
  - Representation of the steps
  - Tracking what is completed and what is left

- Multiple domains
  - Model of intents in each domain
  - Set priorities of tasks in different domains

- Multimodal
  - Represent the capabilities of each modality
  - Include visual modalities in context

- Goal Driven
  - Recognize the user’s goals
  - Select tasks that meet those goals

- Reach to the IoT
  - Model physical context
  - Track patterns
From App capabilities to Discourse requirements

- LUI becomes more complex as well

- **Multi-step tasks**
  - Reference resolution
  - Verbal clues to map steps

- **Multiple domains**
  - Ability to disambiguate
  - Knowing when to clarify

- **Multimodal**
  - Resolving deictic references
  - Resolving/generating verbal

- **Goal Driven**
  - Inferring goals from tasks
  - Recognizing goal changes

- **Reach into the IoT**
  - Semantic model of what things can do
  - Variable degrees of clarification
Discover Deis:

Web App to Navigate Brandeis
Discover Deis

# Tasks
Multitasking
Domains
Goal Driven
Multimodal

# Devices
“Reach” / IoT

Personalization
Mxd Initiative
Self aware

# Participants

- Give a tour while keeping to a schedule → Build task sequence on the fly
- Remember where I parked my car → Storing personal information. Multimodal.
- Recommend events → Temporal model, personalization
- Show a map of where the car is parked while giving directions → Coordination of visual and language
- Help new students get to their classes → log in and access student’s schedule; Semantic model of courses and descriptions “my math class”
- Point out landmarks related to students interests → mixed initiative
Metrics for evaluation?

- Numbers based evaluation is possible
  - Ask the research community about the dialog challenges

- Have to first focus on understanding the capabilities we want in our apps
  - Application driven measures of complexity

- Next, project capabilities to the requirements of the dialog manager

- Finally, drive implementation and evaluation from use cases
  - Remember last year “Wait, do I have enough in my budget for that?”
Thanks!

- The AVIOS Advanced Dialog Group
  - Emmet Coin
  - Marie Meteer
  - John Tadlock
  - K.W. (Bill) Scholz
  - Lorin Wilde

- The students in the Brandeis Discourse and Dialog seminar
  - Tuan Do, Swini Garimella, Jessica Huynh, Alex Luu, Orion Montoya, Hannah Provenza, Keigh Rim,

- See me if you're hiring! Summer interns available.