Speech Processing 11-492/18-492

Spoken Dialog Systems
Conversing with machines
**Spoken Dialog Systems**

- **Not just ASR bolted onto TTS**
- **Different styles of interaction**
  - Question/response systems
  - Mixed initiative systems
  - “How May I Help You?” open questions
  - True conversational machine-human interaction
Introduction

Building simple dialog systems

VoiceXML
  • A language for writing systems

Beyond tree-based systems

Beyond spoken language

Non-task-oriented systems

Real-world deployment considerations
SDS Applications

- **Information giving/request**
  - Flights, buses, stocks and weather
  - Driving directions
  - Answer questions, news

- **Transactional**
  - Reply your email
  - Credit card and bank enquiries, product purchase

- **Maintenance**
  - Technical support
  - Customer service
Entertainment
  • Game characters (NPC), toys, robots

Tutoring
  • Math, science
  • Language learning

Health care
  • Depression screening
  • Aphasia therapy
Dialog Types

- **System initiative**
  - Form-filling paradigm
  - Can switch language models at each turn
  - Can “know” which is likely to be said

- **Mixed initiative**
  - Users can go where they like
  - System or user can lead the discussion

- **Classifying:**
  - Users can say what they like
  - But really only “N” operations possible
  - E.g. AT&T? “How may I help you?”
System Initiative

- **Most common**
  - Machine controls the call
  - Few choices in the dialog
- **Simple form filling:**
  - What is your bank account number
- **Advantages:**
  - You know what users will say (sort of)
  - Hard for user to get confused
  - Hard for system to get confused
  - Easy to build
- **Disadvantages:**
  - Limited flexibility in interaction
  - Fixed dialog structure
- **Most reliable, but many turns**
Let’s Go Bus Information
- 412 268 3526 (Anytime)
- Provides bus information for Pittsburgh

Tell Me
- Company getting others to build systems
- Stocks, weather, entertainment
Mixed Initiative

- **User or system takes initiative**
  - More interesting dialogs
  - “jump” through different parts of dialog state

- **Advantages**
  - More realistic dialog
  - Can do more complex tasks

- **Disadvantages**
  - Can get confusing
  - Can miss important parts
Classification Dialogs

- **Sort out from N things**
  - User says “anything” and system directs them
  - Receptionist
    - I have a problem with my bill
    - What’s the area code for Miami
    - Did you know I can see the beach from here

- **Advantages**
  - (Apparently) complex understanding
  - Solves a very common task

- **Disadvantages**
  - Actually quite restrictive
  - Needs data to train from
  - Needs to be updated
Beyond Telephones

- **Telematics**
  - Voice communication in cars
  - CPS, music selection etc

- **Web-based dialog systems**

- **Robot Interaction**
  - Robot-robot and robot-human interaction

- **Animated talking head**
  - Non-player characters – web agents

- **Speech to Speech translation**

- **CMU Skylar: integrating many dialog systems**
Team Talk

- **Using speech to control multiple robots**
  - Robots have names and distinct voices
  - They report to each other and to you in voice
Other SDS

- **Microsoft: Situated Interaction**
  - *Talking Head that follows you*

- **CMU SV: Aidas**
  - *Restaurant recommendations in situ*
True conversation

- Requires more than just speech
  - Non-verbal noises: laughing, er, um, etc
  - Eye gaze
  - Proper timing (not waiting 500ms before speaker)
  - Back-channeling
  - Movement
  - Talking about nothing
Entrance to NSH

- Keyboard (no ASR)
- TTS, face, movement
- Range finder to detect people
- Significant background character

Mostly talks about nothing
Personal Intelligent Systems

- Example: Apple Siri, Google Now, Microsoft Cortana, Amazon Echo, etc.
- Hub of all applications
- Extendable
- Personalization
- Cross-Language
- Cross-Cultural
- Future: interface-> true companion
Speech Processing 11-492/18-492

Spoken Dialog Systems
SDS components
More than just ASR and TTS

- Recognition
- Language understanding
- Manipulation of utterances
- Generation of new information
- Text generation
- Synthesis
SDS Architecture

ASR → Language Understanding → Dialog Manager

Synthesis → Language Generation → Dialog Manager
Language Understanding
- From words to structure

Dialog Manager
- State of dialog (who is talking)
- Direction of dialog (what next)
- References, user profile etc
- Interaction of database/internet

Language Generation
- From structure to words
 Parsing of SPEECH not TEXT

- *Eh, I wanna go, wanna go to Boston tomorrow*
- *If its not too much trouble I’d be very grateful if one might be able to aid me in arranging my travel arrangements to Boston, Logan airport, at sometime tomorrow morning, thank you.*
- *Boston, tomorrow*
"I wanna go to Boston, tomorrow"
- Destination: BOS
- Departure: 20081028, AM
- Airline: unspecified
- Special: unspecified

Convert speech to structure
- Sufficient for further processing/query
User:
find a cheap eating place for Taiwanese food

Intelligent Agent:
Cheap Taiwanese eating places include Din Tai Fung, Boiling Point, etc. What do you want to choose? I can help you go there.
User

find a cheap eating place for taiwanese food

Intelligent Agent

seeking

price

AMOD

food

NN

target

PREP_FOR
User

find a cheap eating place for Taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User

find a cheap eating place for taiwanese food

Ontology Induction (semantic slot)

Intelligent Agent

User
find a cheap eating place for Taiwanese food
User

find a cheap eating place for taiwanese food

Intelligent Agent

seeking="find"
target="eating place"
price="cheap"
food="taiwanese"
User

seeking = "find"
target = "eating place"
price = "cheap"
food = "taiwanese"

Intelligent Agent

SDS Process
find a cheap eating place for taiwanese food

price

AMOD

food

target

PREP_FOR

Semantic Decoding
can i have a cheap restaurant

Frame: capability

Frame: expensiveness

Frame: locale by use

slot candidate
 Parsing vs Language Model

- **Language Model**
  - Model what actually gets said

- **Parsing**
  - Extract the information you want

- **Models *can* be shared**
  - Only accept things in the grammar
  - Can be over limiting
Neural Networks for SLU

- **RNN for Slot Filling**

  Mesnil et al. 2013

<table>
<thead>
<tr>
<th>Input (words)</th>
<th>show</th>
<th>flights</th>
<th>from</th>
<th>Boston</th>
<th>to</th>
<th>New</th>
<th>York</th>
<th>today</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (labels)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B-dept</td>
<td>0</td>
<td>B-arr</td>
<td>I-arr</td>
<td>B-date</td>
</tr>
</tbody>
</table>

- **Step 1:** word embedding
- **Step 2:** short-term dependencies capturing
- **Step 3:** long-term dependencies capturing
- **Step 4:** different types of neural architecture

http://deeplearning.net/tutorial/rnnslu.html#rnnslu
Interactive Learning for SLU

Luis: Interactive machine learning for language understanding

Advantages:
- Non-expert could add in knowledge in feature engineering
- Active-learning reduces heavy labeling

https://www.luis.ai/

Williams et al. 2016
**Maintain state**

- *Where are we in the dialog*
- *Whose turn is it*
  - Waiting for speaker
  - Waiting for database query (stall user)
- *Deal with barge-in*