

Spoken Human Robot Interaction

Dialogue Management



SAPIENZA
UNIVERSITÀ DI ROMA

Andrea Vanzo, Daniele Nardi
lastname@dis.uniroma1.it
Department of Computer, Control, and Management Engineering
"Sapienza" University of Rome, Italy

Artificial Intelligence
AY 2018/19

Spoken Dialogue System (SDS)

- **Spoken Dialogue Systems** are intelligent agents able to help users to finish tasks more efficiently through spoken interactions
- **SDSs** are being integrated into various devices



SHRI - Dialogue Management mercoledì 12 dicembre 2018 2

Dialogic Interactions

From question-answering to dialog

- Communication takes place in a sequence of turns
- Each sentence depends (at least) on the previous one, in general the dialogue state influences the flow
- Language in dialog is more "implicit" and unstructured
- Disfluencies can be relevant (including other sound features)
- Text-To-Speech is required

In Robotics there is also a strong connection with other HRI components (e.g. facial expression, gestures)

Complex architecture: dialogue manager, SLU, dialogue state tracking, natural language generation, ...

SHRI - Dialogue Management mercoledì 12 dicembre 2018 3

Types of SDS

Conversational agents

- Personal Assistants
- Help user to achieve manifold tasks

(Chit-chat) chatbots

- User entertainers
- No specific goal
- Focus is on producing coherent responses

Task-oriented SDSs

- Personal Assistants
- Help user to achieve a specific tasks

SHRI - Dialogue Management mercoledì 12 dicembre 2018 4

Chatbots

Mimic the unstructured dialogical nature of human-human interaction

Rule-based: pattern/response model

- ELIZA (Rogerian psychology)
- PARRY (keep an emotional model of the bot, and expresses emotions in the text, like anger, fear)

Although ELIZA fooled several people, PARRY was the first bot to pass the Turing test

Corpus-based Chatbots

They rely on a corpus of h-h or h-m conversations

- **IR-based:** the next **turn** is selected through an Information Retrieval task, by computing the most similar sentence to the user request
 - Query = user sentence
 - Documents to be retrieved = possible responses
- **Transduction models:** try to find the best transformation of the query (aka of ML ELIZA)

Task-oriented SDSs:

Based on:

- **Domain ontology:** a characterization of the user intensions that can be dealt with
- **Frames:** structured representations of the semantics associated to each intension
- **Slots** (frame elements): containers for the **values** that need to be filled in to achieve the intension

SDSs: Process structure

First, recall that the ASR returns the N-best list

1. Dialogue Act (preliminary)
2. Domain classification
3. Intent determination
4. Slot filling

Dialogues can be system-Initiative/**mixed Initiative**

In Robotics a special case is **Symbiotic Autonomy**, where the tobot asks a user for help

Speech Act (Dialogue Act)

- Represent what type of action a user utterance is making

Dialogue Act	Example
STATEMENT	<i>Me, I'm in the legal department.</i>
WH-QUESTION	<i>What did you wear to work today?</i>
YES-NO-QUESTION	<i>Do you have to have any special training?</i>
...	...

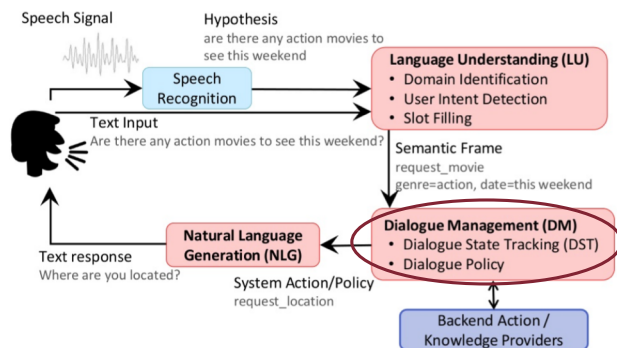
- No standard set recognized by the community...
- ...though the most used is the one defined in Stolcke et al. 2006 (42 DAs)

Stolcke, Andreas, et al. "Dialogue act modeling for automatic tagging and recognition of conversational speech." *Dialogue* 26.3 (2006).

Symbiotic Autonomy: types of dialogic interactions

- Dialogue-based Human Augmented Mapping*
 - Instruct the robot about the environment
- Dialogue-based Task learning*
 - Instruct the robot for new actions to perform
- Reasoning about the environment*
 - E.g. mismatch between semantic map and real world
- Management of robot's self-awareness*
 - E.g. Taking action with busy tray/manipulator
- Dealing with persisting ambiguities*
 - E.g. two objects of the same type are requested

Task-oriented SDSs: Architecture



Task-oriented SDSs: Design Approaches

- Rule-based**
 - Ad-hoc Frame-based architecture (GUS)**
 - Finite State Automata (each state corresponds to a state of the dialog and codes a response -- inflexible)
 - Structured Frame Languages**
 - voiceXML
 - AIML
- Statistical approaches**
 - End-to-end DNN
 - Reinforcement Learning

VoiceXML

Markup Language that is representative of a class of industrial SDS frame-based specification tools

- A VoiceXML document contains a set of dialogs, each a menu or a form.
- A **form** is a frame, whose slots are called **fields**.
- Allows for a grammar specification to parse the input.

Richard Wallace. (2003). *The Elements of AIML Style*.

SHRI - Dialogue Management

mercoledì 12 dicembre
2018

13

VoiceXML

```
<noinput>
I'm sorry, I didn't hear you. <reprompt/>
</noinput>

<nomatch>
I'm sorry, I didn't understand that. <reprompt/>
</nomatch>

<form>
  <block> Welcome to the air travel consultant. </block>
  <field name="origin">
    <prompt> Which city do you want to leave from? </prompt>
    <grammar type="application/x-nuance-gsl">
      [(san francisco) barcelona (new york)]
    </grammar>
    <filled>
      <prompt> OK, from <value expr="origin"/> </prompt>
    </filled>
  </field>
  <field name="destination">
    <prompt> And which city do you want to go to? </prompt>
    <grammar type="application/x-nuance-gsl">
      [(san francisco) barcelona (new york)]
    </grammar>
    <filled>
      <prompt> OK, to <value expr="destination"/> </prompt>
    </filled>
  </field>
  <field name="departdate" type="date">
    <prompt> And what date do you want to leave? </prompt>
    <filled>
      <prompt> OK, on <value expr="departdate"/> </prompt>
    </filled>
  </field>
  <block>
    <prompt> OK, I have you are departing from <value expr="origin"/>
      to <value expr="destination"/> on <value expr="departdate"/>
    </prompt>
    send the info to book a flight...
  </block>
</form>
```

SHRI

AIML--Artificial Intelligence Markup Language

“Stimulus/Response” (S/R) pattern (used in common chatbots)

- **Stimulus** represents what the user may say, and is the input of the Interpreter
- **Response** represents what the user expects as answer, given the corresponding stimulus. It is the output, that can be:
 - A string
 - A system call

Richard Wallace. (2003). *The Elements of AIML Style*.

SHRI - Dialogue Management

mercoledì 12 dicembre
2018

15

AIML Basics

- <aiml>** begins and ends an AIML document
- <category>** defines a unit of knowledge (interaction)
- <pattern>** matches what the user say (stimulus) - RegExp
- <template>** defines the response to the user's input (response)

```
<aiml>
  <category>
    <pattern>HELLO</pattern>
    <template>Hi human!</template>
  </category>
  <category>
    <pattern>GO TO THE KITCHEN</pattern>
    <template>action:GOTO_dest:kitchen</template>
  </category>
</aiml>
```

SHRI - Dialogue Management

mercoledì 12 dicembre
2018

16

Richard Wallace. (2003). *The Elements of AIML Style*.

Dialogue Management via RL

Incrementally Learning Semantic Attributes through Dialogue Interaction

Scenario: **Acquisition of objects' semantic properties**

- Learning dialogue policies to support semantic attributes acquisition
 - Focus is on objects category (e.g., stapler, book, pencil, ...)
- The instruction process is supported by a visual classifier
 - Provides initial guess about the category
- The dialogic interaction for the acquisition of a category consists of a sequence of turns
 - Mixed initiative
- **Goal:** acquire objects' category while minimizing the tutoring cost (trivially, number of turns)

Evaluating SDSs

Performance Metrics:

- Qualitative
 - User satisfaction rating (Questionnaire)

TTS Performance	Was the system easy to understand ?
ASR Performance	Did the system understand what you said?
Task Ease	Was it easy to find the message/flight/train you wanted?
Interaction Pace	Was the pace of interaction with the system appropriate?
User Expertise	Did you know what you could say at each point?
System Response	How often was the system sluggish and slow to reply to you?
Expected Behavior	Did the system work the way you expected it to?
Future Use	Do you think you'd use the system in the future?

- Quantitative (few standards e.g, PARADISE)
 - Task completion success
 - Efficiency cost
 - Quality cost (failure to properly answer)

SDS design

User centered design:

1. Study the user and the task
2. Build simulations and prototypes (Wizard-of-Oz)
3. Iteratively test the design on users

Ethical Issues (key issues in learning approaches)

- Bias due to the training corpus
 - TAY chatbot (MS) removed after 16 hours of interaction, as it started posting racist slurs, conspiracy theories and personal attacks
- Privacy
 - ELIZA dialogues
 - Password given
- Gender equality (why female voices?)

References

Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Chapter 24
<https://web.stanford.edu/~jurafsky/slp3/>

Additional reading

Andrea Vanzo, Jose L. Part, Yanchao Yu, Daniele Nardi, Oliver Lemon. *Incrementally Learning Semantic Attributes through Dialogue Interaction*. *Proc. of AAMAS 2018*. Stockholm, Sweden, July 2018.