



Multimodal Dialogue Systems

Volha Petukhova Saarland University

Summer Semester 2019

Formalities

7 CET

- *1. one presentation* for each participant based on a research paper (40% of final scores)
- 2. active participation in discussions and experiments (20% of final scores)
- 3. Final assignment: 10 pages report (40% of final scores)

4 *CET*

Presentation (50%); oral examination (30%) and active participation (20%) Or

Presentation (60%) and active participation (40%)

Introduction

<u>Multimodal natural-language based dialogue</u> as human-machine interface



Examples

- <u>https://www.youtube.com/watch?v=zIFMq5I</u>
 <u>WVjI</u>
- <u>https://www.youtube.com/watch?v=t7Krn-</u>
 <u>DH3tw</u>
- Non-verbal
- <u>https://www.youtube.com/watch?v=YZizCoO</u>
 <u>ctPo</u>
- <u>https://www.youtube.com/watch?v=1X1vNllf</u>
 <u>0xY</u>

Introduction: dialogue systems

(general architecture)



Input Modules

Automatic speech recognition

- Nuance
- Sphinx (http://cmusphinx.sourceforge.net/)
- Kaldi (<u>http://kaldi.sourceforge.net</u>)
- Google API (https://cloud.google.com/speech/)



Modern sensors

Kinect tracking







SMI eye-tracking glasses





Google glass



Biometrical sensors: MYO, Nexus EXG





Intel RealSense technology https://www.youtube.com/watch?v=uINRC83tITA

Interpretation

- Verbal input
- Non-verbal input
- Fusion

Full-fledged dialogue acts

- Feedback acts (68.5%): positive (65.3%), negative (3.2%)
- . Time Management (24.8%)
- . Turn Management (4.7%)
- Discourse Structuring (2%)

Fusion: roles of non-verbal signals

Articulating semantic content (about 39%):

They are relating to the propositional or referential meaning of an utterance For example deictic gestures:

wording: **Press this** little presentation hand:point.....

pure semantic acts, as a rule do not have a communicative function on their own

Communicative function alteration and specification:

- adjustment of the level of feedback (understanding vs agreement)
- express degree of certainty about the validity of the proposition
- reveal speaker's attitude towards the addressee(-s), towards the content of what he is saying, or towards the actions he is considering to perform
- signal speaker's emotional or cognitive state (Pavelin (2002): *modalizers*)

Segmentation



Speech disfluencies



Multimodal: example

D1: wording: WI	nat's teletext?							
gaze: ave	rted personA p	ersonB						
eyes:	narrow							
posture:	working po	sition						
Auto-FB:	Neg. understanding]						
Turn M.	[Turn assign to A						
B1:								
wording:		Um	It's British thing					
gaze: averted	personD perso	nA personD						
eyes:	es: widen							
lips:	random movements							
<i>posture:</i> bowing	working position							
Auto-FB:	Pos. attention							
Turn M.		Turn take	Turn keep					

Example

B:

wording	: But th	I think re	egardless w	e're we're aim	ing for the un	der sixty	five	
gaze:	person D	person A p	erson A	person D	person A	personC	person	A believes that she understood
posture: A: wording gaze: head posture D:	working positio : person B short singl working position	on A believes that to have the tur le nod shor	B wants n t multīplē nod	A believes continues of s(5)	that B wants t as a speaker 	oU	nder sixty	five okay That's a good start reports about the positive valuation of B's utterance and ffers for further debate
<u>Wording</u>	1:							
Gaze:	person_A	person	В			table		
Head:		short	multiple no	-45)			Chelieve	es that B helieves that
Posture:	working position		D bei	lieves that B w	vants to		p and C b	believes that p is true
C:			contii	nues as a spea	ker			
Wording	1:					Уер		/
Gaze:	person D	person A pers	son B				Þ	person A
Head:						long	nods(2)	
Face:						blink	ing	

Posture: working position

Dialogue contribution processing



Tasks of Dialogue Management

- Dialogue flow control
- Dialogue modeling
- →Dialogue context
- →Dialogue acts
 - Dialogue act decision making
 - Dialogue phenomena:
 - Error handling
 - Initiative and cooperation
 - Adaptivity
 -

Dialogue Management: approaches

• Script-based (state machines)

Sequence of pre-defined steps (dialogue script)

• Frame-based (also: form-filling)

Set of slots to be filled (task template) and corresponding prompts

• Plan-based

Collaborative problem solving

Generic paradigm: Information-State Update

Declarative rules for updating dialogue context

Script-Based DM

- Script describes all possible dialogues
- Typically finite state machine
- .Set of states and transitions

State determines system utterance

User utterance determines transition to next state (deterministic)

- •No recursion! (= no nested sub-dialogues)
- Fixed dialogue script
- .OK for system-driven interaction

Finite State Machine

- <States, Init-State, Alphabet, Transition-function>
- Variants: machines having
 - actions associated with states (Moore machine)
 - actions associated with transitions (Mealy machine)
 - multiple start states
 - transitions conditioned on no input symbol (a null)
 - more than one transition for a given symbol and state (nondeterministic finite state machine)
 - states designated as accepting states (recognizer)
 - etc.

See, e.g., NIST http://www.nist.gov/dads/HTML/finiteStateMachine.html"

FSM-Based Models

U: Elevator? S: Hello. Which floor would you like to go to? Floor_n U: Third floor. S: OK, I am taking you to the third floor. floor n init Floor 1 Welcome Ask_floor floor 1 Е States: ... unknown Init-State: ... Alphabet: ... Transition function: ... Not und

FSM-Based Models



FSM-Based Models





FSM-Based DM: Sum Up

- Advantages
 - □ Fixed prompts can be pre-recorded
 - Speech recognition and input interpretation can be tuned for each state
- Disadvantages
 - Rigid dialogue flow
 - □ Inhibiting user initiative
 - Only suitable for simple tasks
 - In principle can make more flexible, but it quickly gets very complex; modular solutions are possible

CSLU toolkit

<u>http://www.cslu.ogi.edu/toolkit/index.html</u>



Frame-Based DM (Form Filling)

Frame (form): what info should be supplied by user

departure_city ? departure_date ? destination_city ? return_date ?

Dialogue states: which slots are filled

General routines for what system should do next (given which slots are filled)

Frame-Based Models

S: Where do you want to go? U: Paris

S: Where will you travel from? U: From Berlin.

S: When will you travel? U: August 1st. departure_city?departure_date?destination_cityParisreturn_date?

departure_cityBerlindeparture_date1/8/05destination_cityParisreturn_date?

•••

Frame-Based Models

S: What can I do for you? U: I want to fly to Paris

departure_city ? departure_date ? destination_city Paris return_date ?

S: Where will you fly from? U: From Berlin on August 1st.

"Overanswering"

departure_cityBerlindeparture_date1/8/05destination_cityParisreturn_date?

Frame-Based Models

- Strategies for deciding what to do next
 - Next unfilled slot
 - □ Slot-combination weighting
 - Ontology-based coherence
- Options for database lookup
 - Delayed (typically; after certain slots filled)
 - Immediate (can be "expensive" = take time, but enables more helpful system behavior)

VoiceXML

- Digital document standard for specifying interactive media and voice dialogue
- Used for developing audio and voice response applications
- VoiceXML documents are interpreted by a voice browser
- VoiceXML application collects and processes info, and plays back info

VoiceXML

- Main elements of a VoiceXML document
 - Form: basic unit of functionality
 - Field: prompts for and accepts user input
 - Prompt: sequence of audio elements or TTS messages
 - Audio: audio file or TTS message to play
 - Filled: processes input, can pass control to other forms
- Form Interpretation Algorithm
 - Defines how fields in a form are filled in , and how the fill ordering can be modified
- Global event handlers (e.g., for error handling, help)
 - Define behavior when predefined global conditions occur
- Control transfer conditions and subroutine constructs (= specialpurpose programming language)
- new, more expressive standard: State Chart XML



See VoiceXML tutorials

http://www.vocomosoft.com/voicexml_tutorial.htm

Some tools (not exactly VoiceXML, but close) www.wit.ai

https://www.luis.ai/

Frame-Based DM: Sum Up

- Advantages
 - More flexible dialogue
 - Enables some user initiative
- Disadvantages
 - Speech recognition more difficult, because user input less restricted
 - Not every task can be modeled by a frame

Plan-Based DM

- Communication is a joint activity: Agents communicate to establish common ground, agents collaborate to accomplish a task
- Collaborative problem solving by (rational) agents
 - Neither agent can accomplish the task alone
 - Need joint goals and mutual understanding
 - Agents collaborate to establish and achieve their goals
- Agents have knowledge about solving tasks
 - deciding on goals (objectives): adopt, select, defer, abandon, release
 - Forming plans to achieve goals (recipes)
- Automated planning: STRIPS; planning operators: actions, preconditions, post-conditions
 - Executing plans (acting)
 - Revising decisions (re-planning, abandoning goals, etc.)
- Agents reason about beliefs and actions
- Intention recognition

Plan recognition

Given: plan for getting a BA

U: I'll take German 101 fall semester.



Collaborative Planning & Acting

User: Send ambulance one to Parma right away.

(initiate (c-adopt (action (send amb1 Parma))))

(initiate (c-select (action (send amb1 Parma))))

System: OK. [sends ambulance]

(complete (c-adopt (action (send amb1 Parma))))

(complete (c-select (action (send amb1 Parma))))

System: Where should we take the victim once we pick them up?

(initiate (c-adopt (resource (hospital ?x))))

User: Rochester General Hospital.

(continue (c-adopt (resource (hospital RocGen))))

System: OK.

(complete (c-adopt (resource (hospital RocGen))))

[Blaylock et al. 2003]

Plan-based DM

- Advantages
 - Flexibility and adaptivity
 - Any task can be modeled
 - ... the ultimate solution
- Disadvantages
 - Specifying planning operators is as hard as writing dialogue scripts
 - Plan recognition is a hard problem
 - Lots of reasoning needed

Dialogue Modeling as Information State Update





Information State

- Representation of the current state of dialogue
- Used by system to
 - Interpret user's contribution
 - Decide which actions to take
 - Decide what to say
 - Store information (dialogue context representation)
- Utterances update information state
- Approaches to DM differ in how IS is represented, what role it plays, what it contains

ISU Dialogue Modeling

Components:

- a description of the informational components of the IS (aspects of common context, participants, common ground, linguistic and intensional structure, commitments, beliefs, intentions, user model...)
- their formal representation (e.g. lists, sets, typed feature structures, DRSs, propositions, modal operators, etc.)
- set of dialogue acts (DAs) triggering the update of the IS
- set of update rules governing the IS updates given various conditions of current IS and performed DAs (e.g. set of selection rules that license choosing a particular DM to perform given IS)

- a control strategy to decide which update rule(s) to select at a given point in the dialogue (e.g. "pick first that applies", game theory, statistical methods)

IS Update Rules

Describe possible transitions from one information

state to the next

If <conditions-on-IS-values> then <changes-to-IS-values>



Conditions: when a rule is applicable

Effects: how the IS changes

State Machine Model as ISU

- IS: current-state; input
- Update rules:

If [state] & [input] then [output]; [next-state]

Frame-Based Model as ISU

- IS: task-frame; user's move; system move
- Update rules: e.g.,

If [user move = slot X value V] then [fill X with V] If <conditions-on-frame-values> then <ask-slot-value Y>

Decision about next system move is also a rule

ISU-Based Dialogue Modeling

- Task- vs. Dialogue-Structure
 - Task --> dialogue
 - But, dialogue does not have to follow task (execution) structure
- "Dialogue planning": creating an agenda
 - Task model fills agenda with task-related goals
 - Dialogue manager can add more goals, e.g., for grounding
- Some approaches:
 - QUD-based: Godis (TRINDI, SIRIDUS)
 - Obligation-based: Edis (TRINDI)
 - Agent-based: collaborative problem solving: TALK

Chatbots

A chatbot is a conversational agent that interacts with users using natural language.

·First chatbot - ELIZA (Weizenbaum 1966), which emulated a psychotherapist:<u>http://nlp-addiction.com/eliza/</u>

· ALICE is a chatbot: **ALICE System** <u>http://www.alicebot.org/about.html</u>

ALICE: the Artificial Linguistic Internet Computer Entity; a software robot that you can chat with using natural language.

ALICE language knowledge is stored in **AIML** files.

AIML: The Artificial Intelligence Mark up Language.

Alice

•Topics : each Topic file contains a list of categories

- Categories: contain
- Pattern: to match with user input
- > Template: represents ALICE output

```
< aiml version="1.0" >
< topic name=" the topic" >
<category>
<pattern>PATTERN</pattern>
```

```
<template>Template</template></category>
```

</topic>

AIML

•CATEGORIES (Basic unit of knowledge)

```
»<category>
<pattern>HELLO</pattern>
<template>Hi there!</template>
</category>
```

Consists of: Input Question, Output Answer, [Context]

•Pattern = Initial question (a.k.a. "Stimulus")

•Template = Answer (a.k.a. "Response")

•Context = Optional, "that" or "topic"

•Consists only of words, spaces and wildcards _ and *

•Words have letters and/or numerals, space separated

Yak >> AIML files

If you're new to Pandorabots and <u>AIML</u> you should first try customizing your pandorabot by changing some of its <u>properties</u> or by providing your own custom responses with the <u>training</u> interface.

For more advanced botmasters, this page allows you to download, modify and upload the AIML files for your pandorabot directly.

The tables below show all the AIML files for this pandorabot. To view or edit a file, click on its name.

Filename	Size	Last Modified		Browse	Download	Library	Active	Delete
<u>AI.aiml</u>	32,931	09/15/2003 06:05:17	РМ	Browse	Download	aaa-09-2003	•	
ALICE.aiml	36,385	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003	•	
<u>Astrology.aiml</u>	1,227	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
Atomic.aiml	427,931	09/15/2003 06:05:17	ΡМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
Biography.aiml	84,535	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
<u>Bot.aiml</u>	415,844	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
<u>Botmaster.aiml</u>	5,350	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
<u>Client.aiml</u>	138,096	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
<u>Computers.aiml</u>	21,925	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
<u>Default.aiml</u>	623,077	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
Drugs.aiml	11,905	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
Emotion.aiml	15,784	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
Food.aiml	4,901	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
<u>Geography.aiml</u>	167,418	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
<u>History.aiml</u>	1,389	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
<u>Humor.aiml</u>	1,215	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
Inquiry.aiml	3,298	09/15/2003 06:05:17	РМ	Browse	<u>Download</u>	aaa-09-2003		
Interjection.aiml	2,613	09/15/2003 06:05:17	РМ	<u>Browse</u>	<u>Download</u>	aaa-09-2003		
<u>IU.aiml</u>	1,369	09/15/2003 06:05:17	РM	Browse	Download	aaa-09-2003	~	
Knowledge.aiml	266,551	09/15/2003 06:05:17	PM	<u>Browse</u>	<u>Download</u>	aaa-09-2003		

Let's talk to Alice

Virtual Human toolkit: NPCEditor

\$			NPC Editor	- C:\Users\Ram	\Desktop\p	oroject\ł	aydntest1inpi	rogress.plist				×
File Edit Dat	tabase Serv	er View Window	Help									
Utterances s	Settings Peo	ple Classifiers Co	nversations Dialog Manager Chat									
Questio	ons		Filter	Q	Answe	ers			1	Filter	Q	
ID	Score	Clarity A Clarit	v O Text	# Domai	ID	#	Score	Text	Type	External ID	Modified	¢
137		0.0	0.0 what are you famous for?	4 0 0 Ha	5	6	9	I am dead.	- // A	Anybody-9	19/06/14 21	
138		0.0	0.0 Why are you particularly famous for?	3 00 Ha	<u> </u>	7	2	Yes, I am dead.	A	Anybody-10	20/06/14 20	. ^
139		0.0	0.0 can you tell me what you are famous for?	400	5	8	3	I died on 31 May 1809.	4	Anybody-12	19/06/14 21	
140		0.0	0.0 do you know for what you are famous for?	300	5	9	9	I no longer live.	A	Anybody-11	19/06/14 21	
141	6	0.0	0.0 What did you do for living?	300	6	0	2	I am not alive	4	Anybody-212	2 23/09/14 23	
142	6	0.0	0.0 can you tell me what did you do for living?	300	6	1	4	She is not alive	A	Anybody-213	3 23/09/14 23	
143	6	0.0	0.0 do vou know how did vou manage vour living?	3000	6	2	2	They are not alive	4	hybody-215	5 23/09/14 23	
144	-	. 0.0	0.0 Are you a famous composer?	200	6	3	2	I was not considered a child prodigy	. А	Anybody-177	7 04/07/14 00	
145		0.0	0.0 can you tell me if you are a famous composer?	20011	6	4	28	I was a prominent and prolific	A	hybody-13	19/06/14 21	
146		0.0	0.0 Were you a famous musician?	2 0 0 H a	6	5	12	I was a famous composer.	4	Anybody-14	19/06/14 21	
147		0.0	0.0 can you tell me if you are a famous musician?	3 0 0 H a	6	6	11	I was a music composer. I have	A	hybody-15	19/06/14 21	
148	6	0.0	0.0 But what did you do when you are alive?	3001	0 6	7	13	I composed music.	A	Anybody-16	30/07/14 12	
149	6	0.0	0.0 What is your profession?	40	6	8	23	I composed classical music.	A	hybody-18	20/06/14 15	
150		. 0.0	0.0 can you tell me your profession?	3001	6	9	8	I was born on 31 March 1732.	A	Anybody-38	19/06/14 21	
151	6	0.0	0.0 What is your occupation?	40 / 4	7	0	4	I was 77 years old when I died.	4	Anybody-119	21/06/14 14	
152	-	. 0.0	0.0 can you tell me your occupation?	3 0 0 Ha	/ 7	1	15	I was born in 1732 and died in 1809.	A	Anybody-59	19/06/14 21	•
153		0.0	0.0 Did you graduate from university?	2 OG Ha	7	n .	2	I was born 282 years and		mybody-60	10/06/14 21	× 1
154		0.0	0.0 What is your education level?	2 OG H	<						>	•
155		0.0	0.0 could you tell me your education level?	2 OG 14	Scores for:	OG Ha	/dn	1 of 252	Ad	ld.	Demove	
156		0.0	0.0 do vou know what vou studied?	2 OG Ha	scores for.	401na	yun	• 101252	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	iu i	Remove	
157		0.0	0.0 Were you a child prodigy?	1 0 0 Ha	External ID		v-16					
158		0.0	0.0 can you tell me if you are a child prodigy?	1 0 0 Ha	External 10	Anybot	y-10					_
159		0.0	0.0 When did you live?	2 OG Ha	Domain	n: QG_Ha	ydn	~	Type:	No Value		
160		0.0	0.0 can you tall no when did you live?	200 -	×11				type.	NO Value		~
<				>					Speaker:	Haydn		~
Scores for:	QG_Haydn	~	0 of 673 Add	Remove					Addressee:	User		~
External ID:	No Selection	1							Topic:	No Value		~
Domain:	No Selection	n v	Addressee: No Sele	ection					Sound:	No Value		\checkmark
			Topic: No Sele	ection					To Do:	No Value		~
			To Do: Ma Selá	ection	I comp	osed m	usic.					