



CSCI 544, Lecture 9: Dialogue (cont.); Experiment design; Annotation

Ron Artstein

2022-09-20

These notes are not comprehensive, and do not cover the entire lecture. They are provided as an aid to students, but are not a replacement for attending class, participating in the discussion, and taking notes. Any distribution, posting or publication of these notes outside of class (for example, on a public web site) requires my prior approval.

Administrative notes: deadlines



Written Assignment Peer Grading due September 22

- So far, about 40% of the students have completed the grading

Coding Assignment 2 due September 27

Project:

Due Date	Task
September 20	Form project teams (52 teams)
September 20–29	Initial discussion with TA
October 4	Project proposal
November 3	Project status report
Nov 29/Dec 1	Poster presentations (in class)
December 1	Final report
December 3	Self-evaluation and peer grading

Pick your TA!

Administrative notes: missed class



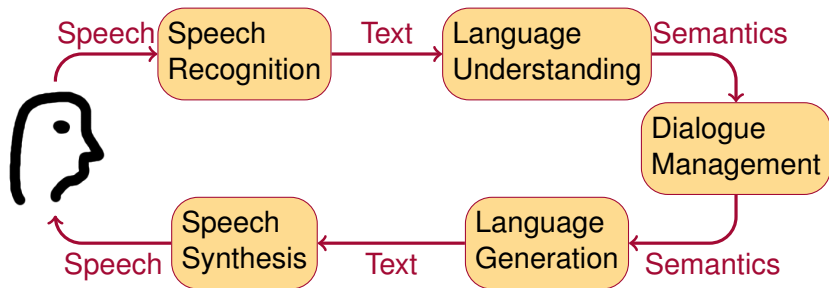
From the [course syllabus](#):

- Students who are absent from class for any reason must make up the materials themselves, and must submit their assignments on time.
 - We will make the presentation slides available on Blackboard, and will endeavor to provide lectures both in-person and on Zoom.
- 👉 Responsibility is **on the student**.

Also on the [course syllabus](#):

- There will be a small number of extra quizzes to allow students to make up for quizzes missed due to late registration, illness, or other reasons. Make-up quizzes will be given after all of the regular quizzes have been taken.

Spoken dialogue systems: prototypical architecture



Information flow, not pipeline

- 👉 Input does not necessarily result in output

Types of dialogue systems



Task/goal-based: Specific, external task

Conversational: Sustain conversation

Most systems have some of both elements
Conversational goals (even without external goal)

Dialogue state tracking



Dialogue State = where the conversation is now

- A set of state variables
- For task-oriented: which slots have been filled

Dialogue System Technology Challenge
State Tracking

Relatively well-defined for simple tasks

Complex tasks might

- share slot values between tasks
- express complex goals in a single utterance
- interleave related tasks

Task lineages: framework



Lee and Stent (2016). Task Lineages: Dialog State Tracking for Flexible Interaction. Sigdial.

Multi-task dialogue (slot-filling)

What is a task? — Book a restaurant, book a ride, ...

- Each task requires some information in order to be executed

Task schema required and optional slots for operationalization

Dialogue act item slot + value

Task frame parse tasks + associated DAI + confidence + ...

Task lineage history of states; maintain in parallel due to ambiguity

Update function extend, then prune (global)

Conversational dialogue systems



Emphasis on maintaining conversation

Additional goals: education, entertainment, connection

New Dimensions in Testimony demo

- Conversational question-answering character
- Primarily reactive (responds to user questions)

Retrieval-based system



Train = linked questions and answers

Questions			Answers		
ID	Text	External ID	ID	Text	
15	I'm really good thank...	AnsDoc	2	right	
16	how are you doing in...	AnsDoc	3	hi how are you	
17	how are you this...	AnsDoc	4	not too bad	
18	so um tell what does it...	AnsDoc	5	well i think it's a...	
19	what is the holocaust	AnsDoc	6	well the holocaust uh...	
20	when you're asked what...	AnsDoc	7	the holo the holocaust is...	
21	when somebody asks...	AnsDoc	8	uh to be jewish is to be...	
22	so what does it mean to...	AnsDoc	9	to be jewish it means...	
23	you experienced being...	AnsDoc	10	a concentration camp...	
24	and if you were asked...	AnsDoc	11	a a death camp was a...	
25	you were also um in the...	AnsDoc	12	a a ghetto was the first...	
26	at the time you wouldn't	AnsDoc	13	well the final solution	

Estimate most likely language model for the response to Q :

$$P(w|Q) \cong \frac{\sum_j \pi_{R_j}(w) \prod_{q \in \text{tok}(Q)} \pi_{S_j}(q)}{\sum_j \prod_{q \in \text{tok}(Q)} \pi_{S_j}(q)}$$

Where: (S_j, R_j) are all linked question-response pairs

π_A is the language model of utterance A

Rank available responses based on distance to estimated LM.

Top-ranked response is the most appropriate.

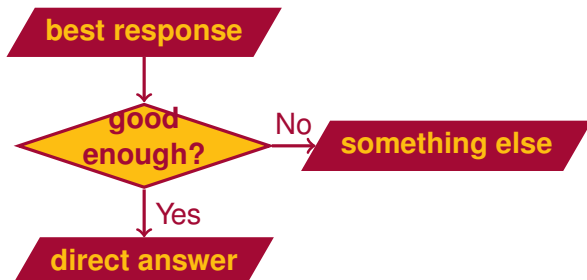


Non-understanding



Direct response not always available

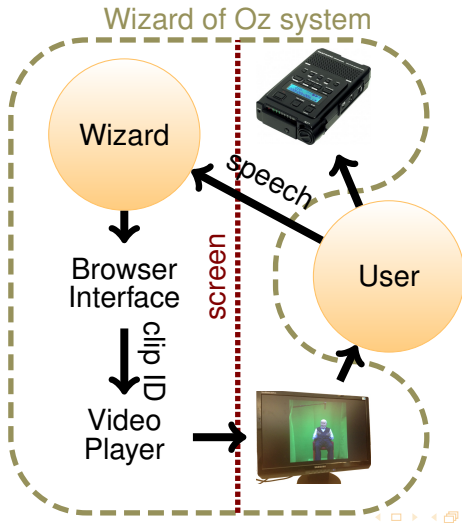
What's your favorite restaurant?



Key to success: good data



Initial system with responses
Collect data in interaction
New responses to fill gaps
More questions for training





Retrieval has limited expressibility

Generation:

- Template
- Grammar
- Sequence-to-sequence (neural)

Conditioned on what?



Typical evaluation of NLP: compare to “gold standard” reference

- Accuracy
- Precision, recall, F-measure

Evaluation of generation: lexical similarity to reference

- Word Error Rate (speech recognition)
- [BLEU](#), METEOR (translation); ROUGE (summarization)
- Multiple references capture lexical variation

Are lexical similarity measures good for dialogue?

- Much more lexical variation in appropriate utterances