

Dialogue and Conversational Agents (part 2)

Ing. R. Tedesco. PhD, AA 19-20

(mostly from: Speech and Language Processing - Jurafsky and Martin)

Outline

A) The Linguistics of Conversation

B) Basic Conversational Agents

- ASR
- NLU
- Generation
- Dialogue Manager

C) Dialogue Manager Design

- Finite State
- Frame-based
- Initiative: User, System, Mixed

D) Information-State

- Dialogue-Act Detection
- Dialogue-Act Generation

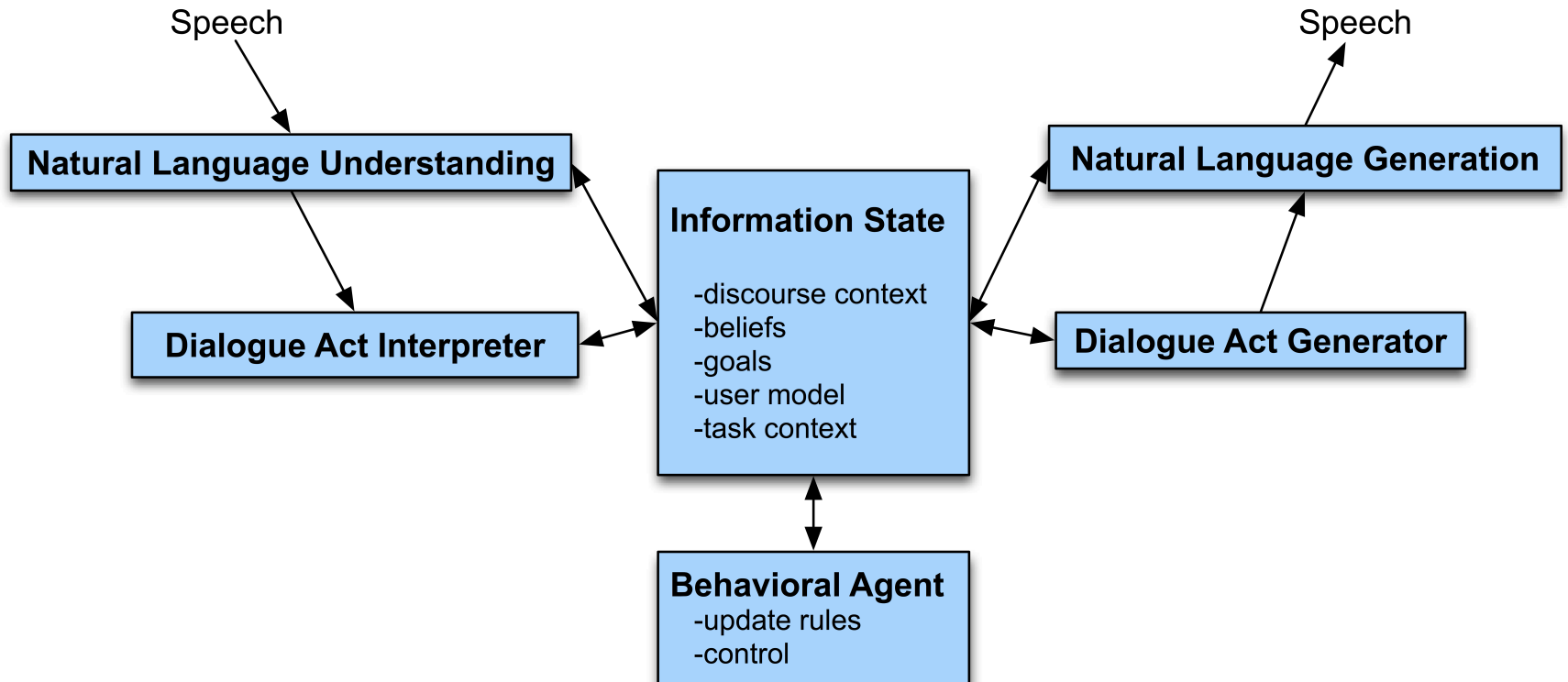
D) Information-State and Dialogue Acts

- If we want a dialogue system to be more than just form-filling
- Conversational agent needs sophisticated models of interpretation and generation
 - In terms of speech acts and grounding
 - Needs more sophisticated representation of dialogue context than just a list of slots

Information-state architecture

- Information state
- Dialogue act interpreter
- Dialogue act generator
- Set of update rules
 - Update dialogue state as acts are interpreted
 - Generate dialogue acts
- Control structure to select which update rules to apply

Information-state



Dialogue acts

- Also called “conversational moves”
- An act with (internal) structure related specifically to its dialogue function
- Incorporates ideas of grounding
- Incorporates other dialogue and conversational functions that Austin and Searle didn't seem interested in

Verbmobil task

- Two-party scheduling dialogues
- Speakers were asked to plan a meeting at some future date
- Data used to design conversational agents which would help with this task
- (cross-language, translating, scheduling assistant)
- Other models exist; e.g.: “DAMSL: Dialog Act Markup in Several Layers”, Allen et al. '97

Verbmobil Dialogue Acts

Act	Example
THANK	thanks
GREET	Hello Dan
INTRODUCE	It's me again
BYE	All right, bye
REQUEST-COMMENT	How does that look?
SUGGEST	June 13th through 17th
REJECT	No, Friday I'm booked all day
ACCEPT	Saturday sounds fine
REQUEST-SUGGEST	What is a good day of the week for you?
INIT	I wanted to make an appointment with you
GIVE_REASON	Because I have meetings all afternoon
FEEDBACK	Okay
DELIBERATE	Let me check my calendar here
CONFIRM	Okay, that would be wonderful
CLARIFY	Okay, do you mean Tuesday the 23rd?

Automatic Interpretation of Dialogue Acts

- How do we automatically identify dialogue acts?
- Given an utterance:
 - Decide whether it is a QUESTION, STATEMENT, SUGGEST, or ACK
- Recognizing **illocutionary force** will be crucial to building a dialogue agent
- Perhaps we can just look at the form of the utterance to decide (locutionary force)?

Can we just use the surface syntactic form?

- YES-NO-Q' s have auxiliary-before-subject syntax:
 - *Will breakfast be served on USAir 1557?*
- STATEMENTS have declarative syntax:
 - *I don't care about lunch*
- COMMAND' s have imperative syntax:
 - *Show me flights from Milwaukee to Orlando on Thursday night*

Surface form \neq speech act type

	Locutionary Force	Illocutionary Force
<i>Can I have the rest of your sandwich?</i>	Question	Request
<i>I want the rest of your sandwich</i>	Declarative	Request
<i>Give me your sandwich!</i>	Imperative	Request

Dialogue Act ambiguity

- *Can you give me a list of the flights from Atlanta to Boston?*
 - This looks like an INFO-REQUEST.
 - If so, the answer is:
 - YES.
 - But really it's a DIRECTIVE or REQUEST, a polite form of: *Please give me a list of the flights...*
- What looks like a QUESTION can be a REQUEST

Dialogue Act ambiguity

- Similarly, what looks like a STATEMENT can be a QUESTION:

Us	OPEN-OPTION	<i>I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next</i>
Ag	HOLD	<i>OK uh let me pull up your profile and I'll be right with you here. [pause]</i>
Ag	CHECK	<i>And you said you wanted to travel next week?</i>
Us	ACCEPT	<i>Uh yes.</i>

Indirect speech acts

- Utterances which use a surface statement to ask a question
- Utterances which use a surface question to issue a request

Dialog Act interpretation as statistical classification

- Lots of clues in each sentence that can tell us which DA it is:
- Words and Collocations:
 - *Please* or *would you*: good cue for REQUEST
 - *Are you*: good cue for INFO-REQUEST
- Prosody:
 - Rising pitch is a good cue for INFO-REQUEST
 - Loudness/stress can help distinguish *yeah*/AGREEMENT from *yeah*/BACKCHANNEL
- Conversational Structure
 - *Yeah* following a proposal is probably AGREEMENT; *yeah* following an INFORM probably a BACKCHANNEL

Statistical classifier model of dialogue act interpretation

- Our goal is to decide for each sentence what dialogue act it is
- This is a **classification task** (we are making a 1-of-N classification decision for each sentence)
- With **N** classes (= number of dialog acts).
- Three probabilistic models corresponding to the 3 kinds of cues from the input sentence.
 1. Conversational Structure: Probability of one dialogue act following another $P(\text{Answer} \mid \text{Question})$
 2. Words and Syntax: Probability of a sequence of words given a dialogue act: $P(\text{"do you"} \mid \text{Question})$
 3. Prosody: probability of prosodic features given a dialogue act : $P(\text{[pitch rises at end of sentence]} \mid \text{Question})$
- It seems an HMM...
 - 1 is the transition probability
 - 2 and 3 are emission probabilities

An example of dialogue act detection: Correction Detection

- Despite all these clever confirmation/rejection strategies, dialogue systems still make mistakes (Surprise!)
- If system misrecognizes an utterance, and either
 - Rejects
 - Via confirmation, displays its misunderstanding
- Then user has a chance to make a correction
 - Repeat themselves
 - Rephrasing
 - Saying “no” to the confirmation question.

Corrections

- Unfortunately, corrections are harder to recognize than normal sentences, for ASR!
 - Swerts et al (2000): corrections misrecognized twice as often (in terms of WER) as non-corrections!!!
- Why?
 - Prosody seems to be largest factor: hyperarticulation
 - English Example from Liz Shriberg
 - “NO, I am DE-PAR-TING from Jacksonville”

Machine learning to detect user corrections

- Useful features:
 - Lexical information (words “no”, “correction”, “I don’ t”, swear words)
 - Prosodic features (various increases in F0 range, pause duration, and word duration that correlate with hyperarticulation)
 - Length
 - ASR confidence
 - Language Model probability
 - Various dialogue features (repetition)

Generating Dialogue Acts

- For two simple cases:
 - Confirmation
 - Rejection

Confirmation

- Another reason for grounding
- **Errors:** Speech is a pretty errorful channel
 - Even for humans; so they use grounding to **confirm** that they heard correctly
- ASR is way worse than humans!
- So dialogue systems need to do even more grounding and confirmation than humans

Explicit confirmation

- S: *Which city do you want to leave from?*
- U: *Baltimore*
- S: *Do you want to leave from Baltimore?*
- U: *Yes*

Implicit confirmation: display

- U: *I'd like to travel to Berlin*
- S: *When do you want to travel to Berlin?*
- U: *Hi I'd like to fly to Seattle Tuesday morning*
- S: *Traveling to Seattle on Tuesday, August eleventh in the morning. Your name?*

Implicit vs. Explicit

- Complementary strengths
- **Explicit:** easier for users to correct systems' s mistakes (can just say “no”)
 - But explicit is cumbersome and long
- **Implicit:** much more natural, quicker, simpler
 - If system guesses right...

Implicit and Explicit

- Early systems: all-implicit or all-explicit
- Modern systems: adaptive
- How to decide?
 - ASR system can give **confidence metric**
 - This expresses how convinced system is of its transcription of the speech
 - If high confidence, use implicit confirmation
 - If low confidence, use explicit confirmation

Rejection

“I’m sorry, I didn’t understand that.”

- Reject when:
 - ASR confidence is too low
 - Best interpretation is semantically ill-formed
- Might have four-tiered level of confidence:
 - Below confidence threshold → reject
 - Above threshold → explicit confirmation
 - If even higher → implicit confirmation
 - Even higher → no confirmation