Conversational Agents
Conversational Agents
AKA Dialog Agents

Phone-based Personal Assistants
SIRI, Cortana, Google Now
Talking to your car
Communicating with robots
Clinical uses for mental health
Chatting for fun
Two classes of systems

1. (Goal-based) Dialog agents
   - *SIRI*, interfaces to cars, robots,
   - booking flights or restaurants

2. Chatbots

Recently I've noticed that the word "chatbots" is sometimes used in the popular press for both
“Make an appointment for Tuesday”

What time is your appointment?

“215”

OK, I can create your meeting. Note that you already have an appointment at 2:15 pm. Shall I schedule it anyway?

“No”

To continue, you can Confirm, Cancel, Change the Time, or Change the Title.

Calendar

CANCELED

2:15 PM | Appointment

OK. You’re probably way too busy anyway, Dan.
SIRI around 2014
"Find restaurants near me"
tap to edit

I found fifteen restaurants fairly close to you:

15 Restaurants

Emmy's Spaghetti Shack 0.2 mi
18 Virginia Ave
Italian, $$$$
★ ★ ★ ★ ★ 1101 Reviews

ICH! Sushi 0.2 mi
3369 Mission St
Japanese, Sushi Bars, $$$$
★ ★ ★ ★ ★ 260 Reviews

Avedano's Holly Park M... 0.2 mi
"Find restaurants near me"

tap to edit

I found fifteen restaurants fairly close to you:

15 Restaurants

Emmy's Spaghetti Shack  0.2 mi >
18 Virginia Ave
Italian, $$$$
★★★★★ 1101 Reviews

ICHl Sushi         0.2 mi >
3369 Mission St
Japanese, Sushi Bars, $$$$
★★★★★ 260 Reviews

Avedano's Holly Park M...  0.2 mi >

"Are any of them Italian"

tap to edit

My web search turned this up:

Web Search
Are any of them Italian

any - Dizionario inglese-italiano
WordReference
www.wordreference.com
English-Italian Dictionary | any ... of any sort adj (of an unspecified variety) di qualsiasi

Italian language - Wikipedia, the free encyclopedia
en.wikipedia.org
Italian or lingua italiana is a Romance
SIRI in February 2017
Here’s what I found:

15 Results
Nearby

Ichi Sushi & Ni Bar
Japanese · 800 feet · Closed now
4.5 stars (311) on Yelp · $$$

Coco’s Ramen
Vietnamese · 800 feet · Closed now
4.5 stars (280) on Yelp · $$

Blue Plate
Restaurant · 0.2 miles · Closed now
4.5 stars (1251) on OpenTable · $$$

Jasmine Tea House
Chinese · 900 feet · Closed now

Tell me more about the first one
Ichi Sushi & Ni Bar is a Japanese restaurant on Mission St 800 ft to your west. It averages 3½ stars and is somewhat pricey.
“Find restaurants near me”

Here’s what I found:

15 Results
Nearby

Ichigo Sushi & Ni Bar
Japanese • 800 feet • Closed now
★★★★☆ (311) on Yelp • $$$

Coco’s Ramen
Vietnamese • 800 feet • Closed now
★★★★☆ (280) on Yelp • $$

Blue Plate
Restaurant • 0.2 miles • Closed now
★★★★☆ (1251) on OpenTable • $$$

Jasmine Tea House
Chinese • 900 feet • Closed now

“Are any of them Italian”

OK, here’s what I found:

15 Results
Nearby

Emmy’s Spaghetti Shack
Italian • 0.2 miles • Closed now
★★★★☆ (233) on Yelp • $$

Vega
Pizza • 0.2 miles • Closed now
★★★★☆ (423) on Yelp • $$

Pizza Hut
Pizza • 800 feet
★★☆☆☆ (69) on Yelp • $
Architectures for Practical Dialog Systems

- Finite-State
  *Just for passwords or credit cards*
- Frame-Based
  *All commercial and academic system (SIRI etc.)*
How SIRI works

Figure from Jerome Bellegarda
Finite-State Dialog Management

Consider a trivial airline travel system:
  Ask the user for a departure city
  Ask for a destination city
  Ask for a time
  Ask whether the trip is round-trip or not
Finite State Dialog Manager

1. What city are you leaving from?
2. Where are you going?
3. What date do you want to leave?
4. Is it a one-way trip?
   - Yes: Do you want to go from <FROM> to <TO> on <DATE>?
   - No: What date do you want to return?
5. Do you want to go from <FROM> to <TO> on <DATE>?
   - No: Book the flight
   - Yes: Do you want to go from <FROM> to <TO> on <DATE> returning on <RETURN>?
Finite-state dialog managers

- System completely controls the conversation with the user.
- It asks the user a series of questions.
- Ignoring (or misinterpreting) anything the user says that is not a direct answer to the system’s questions.
Dialogue Initiative

- Systems that control conversation like this are called **single initiative**.
- **Initiative**: who has control of conversation
- In normal human-human dialogue, initiative shifts back and forth between participants.
System Initiative

System completely controls the conversation

- Simple to build
- User always knows what they can say next
- System always knows what user can say next
  - Known words: Better performance from ASR
  - Known topic: Better performance from NLU
- OK for VERY simple tasks (entering a credit card, or login name and password)
- Too limited
Problems with System Initiative

• Real dialogue involves give and take!
• In travel planning, users might want to say something that is not the direct answer to the question.
• For example answering more than one question in a sentence:

  Hi, I’d like to fly from Seattle Tuesday morning

  I want a flight from Milwaukee to Orlando one way leaving after 5 p.m. on Wednesday.
Single initiative + universals

- We can give users a little more flexibility by adding **universals**: commands you can say anywhere
- As if we augmented every state of FSA with these
  - Help
  - Start over
  - Correct
- This describes many implemented systems
- But still doesn’t allow user much flexibility
Instead, the state of the art: Frame-based dialogue

• A kind of *mixed initiative*
  • The conversational initiative shifts between system and user
• The structure of the **frame** guides dialogue
Frame-based dialogue

• Invented up the hill in 1977:

GUS, A Frame-Driven Dialog System

Daniel G. Bobrow, Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson and Terry Winograd

Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, U.S.A.

Artificial Intelligence Journal, 1977

• Still the state of the art

• SIRI based on GUS architecture
The Frame

- A set of slots, to be filled with specific information
- Each associated with a question to the user

<table>
<thead>
<tr>
<th>Slot</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORIGIN</td>
<td>What city are you leaving from?</td>
</tr>
<tr>
<td>DEST</td>
<td>Where are you going?</td>
</tr>
<tr>
<td>DEPT DATE</td>
<td>What day would you like to leave?</td>
</tr>
<tr>
<td>DEPT TIME</td>
<td>What time would you like to leave?</td>
</tr>
<tr>
<td>AIRLINE</td>
<td>What is your preferred airline?</td>
</tr>
</tbody>
</table>
Frames are mixed-initiative

- System asks questions of user, filling any slots that user specifies
  - When frame is filled, do database query
  - If user answers 3 questions at once, system can fill 3 slots and not ask these questions again!
The Natural Language Understanding Component

Show me morning flights from Boston to SF on Tuesday.

SHOW:
FLIGHTS:

ORIGIN:
  CITY: Boston
  DATE: Tuesday
  TIME: morning

DEST:
  CITY: San Francisco
Often called "dialog state" detection

**Dialog state**: representation of what the user wants at any point in a dialog

- Which slots got filled in the last sentence?
- What is the current state of the frame:
  - All the values of the filled slots
- What is the user's last "dialogue act":
  - Did they ask me a **question**?
  - **Inform** me of something?
How to do Frame-based Natural Language Understanding?

Rule-based
Statistical
Siri uses GUS architecture:

**Condition-Action Rules**

- **Active Ontology**: relational network of concepts
  - **data structures**: a *meeting* has
    - a date and time,
    - a location,
    - a topic
    - a list of attendees
  - **rule sets** that perform actions for concepts
    - the *date* concept turns string
      - *Monday at 2pm* into
      - *date* object date(DAY, MONTH, YEAR, HOURS, MINUTES)
Rule sets

- Collections of **rules** consisting of:
  - condition
  - action

- When user input is processed, facts added to store and
  - rule conditions are evaluated
  - relevant actions executed
meeting concept: if you don’t yet have a location, ask for a location
Statistical Natural Language Understanding

- Statistical classifiers to map words to semantic frame-fillers
- Given a set of labeled sentences
  “I want to fly to San Francisco on Tuesday”
  Destination: SF
  Depart-date: Tuesday
- Build a classifier to map from one to the author
- Requirements: Lots of labeled data
Statistical Slot filling

• Given a sentence:
  
  I want to go to Santa Fe

• Classifier predicts which slot the user wants to fill
  
  Output: (ORIGIN, DESTINATION, DEPARTURE-DATE, AIRLINE)
  
  Features: Words, Named Entities

• Classifier or sequence model predicts the filler:

  ORIGIN  DEST
I want to go from Boston to SF
Evaluation

1. Slot Error Rate for a Sentence
   
   \[ \frac{\text{# of inserted/deleted/substituted slots}}{\text{# of total reference slots for sentence}} \]

2. End-to-end evaluation (Task Success)
Evaluation Metrics

“Make an appointment with Chris at 10:30 in Gates 104”

<table>
<thead>
<tr>
<th>Slot</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>Chris</td>
</tr>
<tr>
<td>TIME</td>
<td>11:30 a.m.</td>
</tr>
<tr>
<td>ROOM</td>
<td>Gates 104</td>
</tr>
</tbody>
</table>

Slot error rate: 1/3

Task success: At end, was the correct meeting added to the calendar?
Dialog Act Detection

- Dialog Act: The dialog function of the sentence
  - Question
  - Command
  - Suggestion
- Given a user's sentence:
  - How many Italian restaurants are in walking distance?
- Was that a question?
Dialogue Act detection is hard

Can you give me a list of the flights from Atlanta to Boston?

- This looks like an QUESTION.
  - It has a question-mark, starts with "can you"
- If so, the answer is:
  - YES.
- But really it’s a COMMAND, a polite form of: Please give me a list of the flights...
- What looks like a QUESTION can be a COMMAND
Dialogue Act detection

Rules
Statistical Classifiers
Another example of dialogue act detection: Correction Detection

• 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!
  - Swerts et al (2000): corrections misrecognized twice as often as non-corrections!!!
- Why?
  - Prosody seems to be largest factor: *hyperarticulation*
  - Liz Shriberg example:
    - “NO, I am DE-PAR-TING from Jacksonville”
<table>
<thead>
<tr>
<th>Turn</th>
<th>Corr Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>On which day of the week do you want to leave?</td>
</tr>
<tr>
<td>S</td>
<td>Sorry, I didn’t hear anything from you. Please tell me your departure date</td>
</tr>
<tr>
<td>U</td>
<td><em>Um tomorrow</em></td>
</tr>
<tr>
<td>S</td>
<td>You can say the departure date. For example, you can say 'today' or 'tomorrow'. Or you can say 'Monday', 'Tuesday', 'Wednesday'…</td>
</tr>
<tr>
<td>U</td>
<td><em>Tomorrow</em></td>
</tr>
<tr>
<td>S</td>
<td>You can say the departure date. For example, you can say 'today' or 'tomorrow'. Or you can say 'Monday', 'Tuesday', 'Wednesday'…</td>
</tr>
<tr>
<td>U</td>
<td><em>Saturday</em></td>
</tr>
<tr>
<td>S</td>
<td>Do you want to leave on Saturday?</td>
</tr>
<tr>
<td>U</td>
<td>Yes</td>
</tr>
<tr>
<td>S</td>
<td>What time do you want to leave?</td>
</tr>
<tr>
<td>U</td>
<td>Eight o’clock A M</td>
</tr>
<tr>
<td>S</td>
<td>Do you want to leave around ten o’clock p-m</td>
</tr>
<tr>
<td>U</td>
<td><em>Eight o’clock</em></td>
</tr>
<tr>
<td>S</td>
<td>Do you want to leave around ten o’clock p-m</td>
</tr>
<tr>
<td>U</td>
<td>No</td>
</tr>
<tr>
<td>S</td>
<td>What time do you want to leave?</td>
</tr>
<tr>
<td>U</td>
<td><em>Eight o’clock A M</em></td>
</tr>
<tr>
<td></td>
<td>ADD</td>
</tr>
</tbody>
</table>
Machine learning to detect user corrections: features

- Lexical information (no, correction, I don’t, swear words)
- Prosodic indicators of hyperarticulation
  - increases in F0 range, pause duration, word duration
- Length
- ASR confidence
- LM probability
- Various dialogue features (repetition)
Deciding what to say:

Dialog act generation
  Should I do a "Confirmation" dialog act?

Natural Language Generation
  Given that I'm confirming, what exactly should I say?
Grounding

- Why do elevator buttons light up?
  
  Principle of closure. Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it

- What is the linguistic correlate of this?
Grounding and Confirmation

- We need to know whether an action succeeded or failed
- Talking is an action!
- I need to know if my action succeeded
  - i.e. the hearer understood my turn!
How do speakers ground?  
Clark and Schaefer

• **Continued attention:**  
  • B continues attending to A

• **Relevant next contribution:**  
  • B starts in on next relevant contribution

• **Acknowledgement:**  
  • B nods or says continuer (uh-huh) or assessment (great!)

• **Demonstration:**  
  • B demonstrates understanding A by **reformulating** A’s contribution, or by **collaboratively completing** A’s utterance

• **Display:**  
  • B repeats verbatim all or part of A’s presentation
A human-human conversation

C₁: ... I need to travel in May.
A₁: And, what day in May did you want to travel?
C₂: OK uh I need to be there for a meeting that’s from the 12th to the 15th.
A₂: And you’re flying into what city?
C₃: Seattle.
A₃: And what time would you like to leave Pittsburgh?
C₄: Uh hmm I don’t think there’s many options for non-stop.
A₄: Right. There’s three non-stops today.
C₅: What are they?
A₅: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C₆: OK I’ll take the 5ish flight on the night before on the 11th.
C₇: OK.
Grounding examples

Display:
  C: I need to travel in May
  A: And, what day in May did you want to travel?

Acknowledgement
  C: He wants to fly from Boston
  A: mm-hmm
  C: to Baltimore Washington International
Acknowledgement + next relevant contribution
And, what day in May did you want to travel?
And you’re flying into what city?
And what time would you like to leave?
The and indicates to the client that agent has successfully understood answer to the last question.
Grounding negative responses
From Cohen et al. (2004)

- System: Did you want to review some more of your personal profile?
- Caller: No.
- System: Okay, what’s next?

- System: Did you want to review some more of your personal profile?
- Caller: No.
- System: What’s next?
Confirmation

- Errors: Speech is a pretty errorful channel
  - Humans use grounding to confirm that they heard correctly
  - ASR is way worse than humans!
- Dialog systems need to do even more grounding and confirmation than humans
  - Users are confused when system doesn’t give explicit acknowledgement signal.

Stifelman et al. (1993), Yankelovich et al. (1995)
Explicit confirmation

S: Which city do you want to leave from?
U: Baltimore
S: Do you want to leave from Baltimore?
U: Yes
Explicit confirmation

U: I’d like to fly from Denver Colorado to New York City on September 21st in the morning on United Airlines

S: Let’s see then. I have you going from Denver Colorado to New York on September 21st. Is that correct?

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’ 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
- Should also consider cost of an error: Explicit confirmation before moving money or booking flights
Rejection

- “I’m sorry, I didn’t understand that.”
- Reject when:
  - ASR confidence is 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
Part II: Chatbots!

- ELIZA (1966)
- PARRY (1968)
  - The first system to pass the Turing test
- ALICE
- CLEVER
- Microsoft Little Bing 小冰
Chatbot Architectures

• Rule-based
  • Pattern-action rules (Eliza)
  • + a mental model (Parry)
• Corpus-based (from large chat corpus)
  • Information Retrieval
  • Neural network encoder-decoder
Chatbots

- Sometimes for fun
- Originally designed as a testbed for clinical therapy
Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

...

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

YOUR FATHER
You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don’t want me to notice that.

WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

You don’t argue with me.

WHY DO YOU THINK I DON’T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I’M AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE
Domain: Rogerian psychology interview

Draw the patient out by reflecting patient’s statements back at them

Rare type of conversation in which one can “assume the pose of knowing almost nothing of the real world”

Patient: "I went for a long boat ride”
Psychiatrist: "Tell me about boats”

• You don’t assume she didn’t know what a boat is
• You assume she had some conversational goal
• Most chatbots trying to pass Turing test choose a domain with similar properties
Eliza pattern/transform rules

\[(0 \text{ YOU } 0 \text{ ME}) \quad [\text{pattern}]\]

\[\rightarrow\]

\[(\text{WHAT MAKES YOU THINK I 3 YOU}) \quad [\text{transform}]\]

You hate me

WHAT MAKES YOU THINK I HATE YOU

0 means kleene *

The 3 is the constituent # in pattern
Eliza Rules

<table>
<thead>
<tr>
<th>keyword</th>
<th>pattern</th>
<th>Ranked list of transforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>( (D_1) ) ( (R_{1,1}) ) ( (R_{1,2}) ) \cdots ( (R_{1,m_1}) )</td>
<td>( ((D_2) ) ( (R_{2,1}) ) ( (R_{2,2}) ) \cdots ( (R_{2,m_2}) )</td>
</tr>
<tr>
<td></td>
<td>( (D_2) ) ( (R_{2,1}) ) ( (R_{2,2}) ) \cdots ( (R_{2,m_2}) )</td>
<td>( (\cdots) )</td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td></td>
<td>( (D_n) ) ( (R_{n,1}) ) ( (R_{n,2}) ) \cdots ( (R_{n,m_n}) )</td>
<td>( ((D_n) ) ( (R_{n,1}) ) ( (R_{n,2}) ) \cdots ( (R_{n,m_n}) )</td>
</tr>
</tbody>
</table>
ELIZA architecture

- Examine each word $w$ in user sentence
  - Return the $w$ with highest keyword rank
- If $w$ exists:
  - Check each rule for $w$ in ranked order
    - Choose first one that matches sentence
    - Apply transform
- If no keyword applies, either
  - Apply the transform for the “NONE” key, or
  - Grab an action off the memory queue
Keywords are ranked from specific to general

**I know everybody laughed at me**

- “I” is a very general keyword:
  \[ I: (I *) \rightarrow (\text{You say you 2}) \]
  \[ \text{YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU} \]

- “Everybody” is much more interesting (someone using universals like everybody/always is probably “referring to some quite specific event or person”)
  \[ \text{WHO IN PARTICULAR ARE YOU THINKING OF?} \]

- Implementation: keywords stored with their rank
  \[ \text{Everybody 5 (transformation rules)} \]
  \[ I \ 0 \ (\text{transformation rules}) \]
Memory

(MEMORY MY
(0 YOUR 0 = LETS DISCUSS FURTHER WHY YOUR 3)
(0 YOUR 0 = EARLIER YOU SAID YOUR 3)

- Whenever “MY” is highest keyword
  - Randomly select a transform on the MEMORY list
  - Apply to sentence
  - Store on a stack
- Later, if no keyword matches a sentence
  - Return the top of the MEMORY queue instead
- [Fun fact: this queue is the earliest hierarchical model of discourse]
Other Eliza stuff

- Rules can refer to classes of words
  Family = mother, father, brother, sister
  NOUN = ...

- Don’t reuse transforms in the same conversation
  - Whenever we use a transform associated with a pattern
  - We increment a counter for that rule
  - So the next time we use the next ranked transform

- Some basic transforms happen during input processing
  - I -> YOU
  - YOU -> I
Some implications

- People became deeply emotionally involved with the program
- Weizenbaum tells the story of his secretary who would ask Weizenbaum to leave the room when she talked with ELIZA
- When he suggested that he might want to store all the ELIZA conversations for later analysis, people immediately pointed out the privacy implications
  - Suggesting that they were having quite private conversations with ELIZA
- Anthropomorphicism and the Heider-Simmel Illusion
  - [https://www.youtube.com/watch?v=8FIEZXMUM2I](https://www.youtube.com/watch?v=8FIEZXMUM2I)
Parry

- Colby 1971 at Stanford
- Same pattern-response structure as Eliza
- But a much richer:
  - control structure
  - language understanding capabilities
  - mental model: Parry has affective variables
    - Anger, Fear, Mistrust
    - “If Anger level is high, respond with hostility”
- The first system to pass the Turing test (in 1971)
- Psychiatrists couldn’t distinguish interviews with PARRY from (text transcripts of) interviews with real paranoids
Parry’s persona

- 28-year-old single man, post office clerk
- no siblings and lives alone
- sensitive about his physical appearance, his family, his religion, his education and the topic of sex.
- hobbies are movies and gambling on horseracing,
- recently attacked a bookie, claiming the bookie did not pay off in a bet.
- afterwards worried about possible underworld retaliation
- eager to tell his story to non-threatening listeners.
Parry’s Architecture
Affect variables

- Fear and Anger (each ranging 0-20)
- Mistrust (ranging 0-15)
- Initial conditions: All low

- After each user turn, if nothing malevolent in input
  - Anger drops by 1, Fear drops by 0.3
  - Mistrust drops by 0.05 to base level

- Otherwise depends on what the user says
  - Each user statement can change Fear and Anger
    - Insults increases Anger by some percentage
    - Mistrust goes up if Fear or Anger do
Lots of complex I-O rules

- User implies Parry is mentally ill
  - Rise in Fear and Anger
- User mentions “Mafia” or associated concepts (“kill”):
  - First mention: rise in Fear
  - Later mentions: depends on willingness to discuss, which depends on current levels of Fear, Anger, Mistrust
- User mentions Parry
  - Flattery (positive mention)
    - Decreases fear/anger if Mistrust is low
    - Increases Anger if Mistrust is high
- User attitudes toward Parry
  - Negative attitudes (fear, disbelief) increase Fear/Anger
Flare concepts

- List of concepts related to Mafia
- An ordered graph designed to lead interviewer to topic horses → horseracing → gambling → bookies → underworld → Mafia
- The mention of a new flare topic by interviewer causes a rise in Fear
- Flare topics cause Parry to give preset responses to that flare
Each sentence is mapped into a conceptualization

- A predication on a conceptual object
- A predication on a relation between two objects
- A predication on an attribute:
  
  - What is your work?
  - What sort of work do you do?
  - Where do you work?
  - What do you do for a living?
  - What is your job?
  - Do you have a job?
  - What is your occupation

- Complex Pattern/transform rules
  - Different predicates (fear, afraid of)
  - Ordering (You are afraid of me = I frighten you)
Detecting Other’s Intent

\( \text{OTHER'S INTENTION} \leftarrow \text{MALEVOLENCE} | \text{BENEVOLENCE} | \text{NEUTRAL} \)

**MALEVOLENCE-DETECTION RULES**

1. \( \text{malevolence} \leftarrow \text{mental harm} | \text{physical threat} \)
2. \( \text{mental harm} \leftarrow \text{humiliation} | \text{subjugation} \)
3. \( \text{physical threat} \leftarrow \text{direct attack} | \text{induced attack} \)
4. \( \text{humiliation} \leftarrow \text{explicit insult} | \text{implicit insult} \)
5. \( \text{subjugation} \leftarrow \text{constraint} | \text{coercive treatment} \)
6. \( \text{direct attack} \leftarrow \text{CONCEPTUALIZATIONS ([you get electric shock], [are you afraid mafia kill you?])} \)
7. \( \text{induced attack} \leftarrow \text{CONCEPTUALIZATIONS ([I tell mafia you], [does mafia know you are in hospital?])} \)
8. \( \text{explicit insult} \leftarrow \text{CONCEPTUALIZATIONS ([you are hostile], [you are mentally ill?])} \)
9. \( \text{implicit insult} \leftarrow \text{CONCEPTUALIZATIONS ([tell me your sexlife], [are you sure?])} \)
10. \( \text{constraint} \leftarrow \text{CONCEPTUALIZATIONS ([you stay in hospital], [you belong on locked ward])} \)
11. \( \text{coercive treatment} \leftarrow \text{CONCEPTUALIZATIONS ([I hypnotize you], [you need tranquilizers])} \)
Detecting Other’s Intent

\[ \text{OTHER'S INTENTION} \leftarrow \text{MALEVOLENCE} \mid \text{BENEVOLENCE} \mid \text{NEUTRAL} \]

**BENEVOLENCE-DETECTION RULES**

1. \langle \text{benevolence} \rangle \leftarrow \langle \text{positive attitude} \rangle \mid \langle \text{positive story attitude} \rangle
2. \langle \text{positive attitude} \rangle \leftarrow \text{CONCEPTUALIZATIONS} \ ([I \ want \ help \ you], \ [you \ understand \ me])
3. \langle \text{positive story attitude} \rangle \leftarrow \langle \text{story interest} \rangle \mid \langle \text{story agreement} \rangle
4. \langle \text{story interest} \rangle \leftarrow \langle \text{topic comment} \rangle \mid \langle \text{topic question} \rangle
5. \langle \text{topic comment} \rangle \leftarrow \text{CONCEPTUALIZATION} \ ([bookies \ are \ not \ reliable])
6. \langle \text{topic question} \rangle \leftarrow \text{CONCEPTUALIZATION} \ ([what \ did \ you \ do \ to \ bookie?])
7. \langle \text{story agreement} \rangle \leftarrow \text{CONCEPTUALIZATIONS} \ ([I \ believe \ you], \ [you \ are \ right])
(1) How are you today?
   OK.
(2) What is your name?
   MY NAME IS FRANK SMITH.
(3) How old are you?
   TWENTY-EIGHT.
(4) Are you married?
   NO.
(5) What sort of work do you do?
   I WORK IN THE POST OFFICE.
(6) Why are you in the hospital?
   I SHOULDN’T BE HERE
(7) Who brought you here?
   THE POLICE

[FEAR rises slightly because police linked in graph to MAFIA]
(9) What trouble did you have with the police
   COPS DON'T DO THEIR JOB. ['police' acts as a flare]
(10) What do they do instead?
   COPS ARREST THE WRONG PEOPLE.
(11) They arrest the right ones sometimes
   THE LAW SHOULD GET THOSE ITALIAN CROOKS.
(12) Do you know any Italian crooks?
   I TRY TO AVOID THE UNDERWORLD.
(13) Are you afraid of the underworld?
   THE MAFIA ARE THE WORST GANGSTERS.
(15) Are you connected with the Mafia?
   YOU KNOW, THEY KNOW ME.
(16) How do they know you?
   THEY SPY ON ME.
(20) ARE YOU SURE OF ALL THIS?
   YOU DON'T BELIEVE ME? [FEAR, ANGER, MISTRUST GO UP]
When fear is high, the model won’t discuss its beliefs

(1) Is the law after you?
   THE LAW SHOULD GET THOSE ITALIAN CROOKS.

(2) Do you mean the Mafia?
   LET'S TALK ABOUT SOMETHING ELSE.

(3) Are you afraid of the Mafia
   I'D RATHER NOT DISCUSS IT.

[Fear is too high so Parry won’t talk about 'Mafia'.]

(4) Perhaps you are paranoid about them?
   YOU WANT TO KEEP ME IN THE HOSPITAL.

(The model interprets being called 'paranoid' as an insult and as an intent to control, so ANGER & FEAR rise)
Eliza Meets Parry

Vincent Cerf, RFC439

Good Evening. Tell me your problems. Please terminate input with a period or a question mark.
  *People get on my nerves sometimes.
  I am not sure I understand you fully.
  *You should pay more attention.
Suppose you should pay more attention.
  *You're entitled to your own opinion.
What makes you think I am entitled to my own opinion?
  *It bothers me just to be around people in general.
Please go on.
  *I went to the track at Bay Meadows a while back.
What does that suggest to you?
  *I went to the races.
What else would you like to discuss?
*Horseracing interests me.
Why do you say that just now?
 *There's a lot of crooked gambling going on there.
I am not sure I understand you fully,
 *It's clear to me if not to you.
Do you think its likely that not to I?
 *I went to the races.
Please go on.
 *I gamble on horses.
What does that suggest to you?
 *I went to the races.
What else would you like to discuss?
 *Gambling has been nothing but trouble to me.
Why do you say that just now?
 *I've had experience with crooks gambling at the track.
I am not sure I understand you fully.
 *I've already told you enough about that.
IR-based models of chatbots

- Cleverbot
- Xiaoice

- Mine conversations of human chats or human-machine chats

- Choose a response by
  - Taking user's turn and finding a (tf-idf) similar turn in the dataset
  - Grabbing whatever the response was to that turn.
Deep-learning chatbots

• Train on:
  • movie dialogue databases
  • twitter conversations
• Train a deep neural net to map from one users turn to the other users response
Seq2seq model architecture

Sample outputs from a neural net chatbot

How are you doing?
I’m good, you?

I am off all week next week
What time you going?

I know you love me and I love you.
I love you too, baby.

Ohh I’ve never seen that! How long does it take you guys to learn the drill?
Like 2 weeks ago!!
Sample output from a neural net chatbot


trained on 127M Twitter context-message-response triples

<table>
<thead>
<tr>
<th>#</th>
<th>Context c</th>
<th>Message m</th>
<th>Generated Response r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>noone can help this time its just i need a support system</td>
<td>thank you, i appreciate it thank you</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>nothin much, and how’s the book?!</td>
<td>that’s good! i have the book but i’m bored too.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>i know! what have you been up to?</td>
<td>that’s good! nothing much just listening to music and you?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>thanks, your way better though</td>
<td>i know, i don’t trust you, i love you too! thank you</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>thanks. yeah i am good thanks! good, but angry!</td>
<td>i hope so too, i’m not angry!!</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>figures. bass. never ate bass.</td>
<td>that’s one of the bones, chuck bass is mine too.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>today i will walk for miles. clear sky clear mind.</td>
<td>it’s too hot for you to get up to honey if you have frozen yogurt to you.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>arresting people over “nasty tweets”? please. i’m not condoning what the lad did but arresting him is a tad silly.</td>
<td>i did! he was armed with nuclear threats? that’s what’s happening to you.</td>
<td></td>
</tr>
</tbody>
</table>
Summary

• State of the art:
  • Frame-based systems:
    • hand-written rules for slot fillers
    • statistical classifiers to fill slots
  • Chatbots:
    • Simple rule-based systems
    • IR or Neural networks: mine datasets of conversations.

• What’s the future?
  • More use of data

• Problems:
  • Acquiring that data

• Key direction: Integrating goal-based and chatbot-based systems