Dialogue and Understanding

11 May 2021
Where are we?
What have we seen this semester?

Morphological analysis  Question answering
Probabilistic language models  Information extraction
Text classifiers  Machine translation
Word representations
Sequence models
Syntax
Semantics
Further exploration

Some topics of the topics we didn’t have time for:

- Language and vision
- Language and robotics
- NLP for social science
- Natural language inference
- Paraphrase
- Topic modeling
- Structured prediction
- Deeper coverage of neural networks (e.g., transformers)
Recurring themes

The role of machine learning and the evolution of techniques

  Relative frequencies, log-linear models, neural networks

The role of data

  E.g., annotation, bitext

The importance and challenge of evaluation

Useful building blocks

  From high-level views (syntax vs semantics) to low-level tools
  (recurrent neural networks, dynamic programming)
Where are we all?
Just a few years ago

No seq2seq

No attention

No large-scale question-answering/reading comprehension datasets

No TensorFlow or PyTorch

...
NLP methods

Gentlemen, our learner overgeneralizes because the VC-Dimension of our Kernel is too high. Get some experts and minimize the structural risk in a new one. Rework our loss function, make the next kernel stable, unbiased and consider using a soft margin.
NLP data sets

Even for English, most tasks have 100,000 or fewer labeled examples.

And there’s even less data available for other languages.

There are thousands of languages – and hundreds with more than a million native speakers
< 10% of people speak English as their first language

Increasingly popular solution: use more unlabeled data
Machine translation: Pre-training

1. Separately train encoder and decoder as language models:

2. Then train jointly on bilingual data
Machine translation: Self-training

I traveled to Belgium

Translation: Je suis étudiant

MT Model

Je suis étudiant

train

MT Model
Machine translation: Back-translation

Have two machine translation models going in opposite directions: en → fr and fr → en
Huge models and GPT-3
Meet GPT-3. It Has Learned to Code (and Blog and Argue).

The latest natural-language system generates tweets, pens poetry, summarizes emails, answers trivia questions, translates languages and even writes its own computer programs.
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium-sized LSTM</td>
<td>10 M</td>
</tr>
<tr>
<td>ELMo</td>
<td>90 M</td>
</tr>
<tr>
<td>GPT</td>
<td>110 M</td>
</tr>
<tr>
<td>BERT-Large</td>
<td>320 M</td>
</tr>
<tr>
<td>GPT-2</td>
<td>1.5 B</td>
</tr>
<tr>
<td>GPT-3</td>
<td>175 B</td>
</tr>
<tr>
<td>Model</td>
<td>Parameters</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Medium-sized LSTM</td>
<td>10 M</td>
</tr>
<tr>
<td>ELMo</td>
<td>90 M</td>
</tr>
<tr>
<td>GPT</td>
<td>110 M</td>
</tr>
<tr>
<td>BERT-Large</td>
<td>320 M</td>
</tr>
<tr>
<td>GPT-2</td>
<td>1.5 B</td>
</tr>
<tr>
<td>Honey bee brain</td>
<td>~ 1 B synapses</td>
</tr>
<tr>
<td>GPT-3</td>
<td>175 B</td>
</tr>
<tr>
<td>Cat</td>
<td>~ 10 trillion synapses</td>
</tr>
<tr>
<td>Human</td>
<td>~ 100 trillion synapses</td>
</tr>
</tbody>
</table>
GPT-3 and GPT-2 have the same basic architecture – a really big transformer language model

Exception: GPT-3 adds locally banded sparse attention patterns

but GPT-3 is insanely big.

It was trained on about 500 billion tokens from Wikipedia, Common Crawl, and other sources.
These models can even do machine translation. They’re trained on a corpus that’s almost all English, but includes text like this:

”I’m not the cleverest man in the world, but like they say in French: Je ne suis pas un imbecile [I’m not a fool].

In a now-deleted post from Aug. 16, Soheil Eid, Tory candidate in the riding of Joliette, wrote in French: ”Mentez mentez, il en restera toujours quelque chose,” which translates as, ”Lie lie and something will always remain.”

”I hate the word ‘perfume,’” Burr says. ’It’s somewhat better in French: ‘parfum.’

If listened carefully at 29:55, a conversation can be heard between two guys in French: ”-Comment on fait pour aller de l’autre coté? -Quel autre coté?”, which means “- How do you get to the other side? - What side?”

If this sounds like a bit of a stretch, consider this question in French: As-tu aller au cinéma?, or Did you go to the movies?, which literally translates as Have-you to go to movies/theater?

”Brevet Sans Garantie Du Gouvernement”, translated to English: “Patented without government warranty”.
One of the amazing things about these models are their abilities to perform tasks involving language with few or no examples.
There are many potential applications for huge models like GPT-3, but let’s return to where the course began: Talking with computers.
The training data for GPT-3 cut off in October 2019, so I need to tell it what is going on in the world in July 2020.

**Researcher:** It is currently July 2020. For the last four months, the global economy has been disrupted due to the introduction of COVID-19, a novel coronavirus which is highly transmissible. This new coronavirus has created a global pandemic. In response to the pandemic, the governments of the world have instituted a series of lock downs, asking people to stay at home to prevent the spread of the virus. This resulted in the shutdown of many businesses, can you guess which businesses were impacted the most?

**Wise Being:** The businesses that needed to function in a way that allowed humans to interact with each other directly have been shutdown. This includes restaurants, retail stores and most other service based industries. The largest industries that have not been shut down are those that do not require human interaction.
Researcher: How do you think the world will finally defeat the virus?

Wise Being: Well, most of the world is not really trying to stop the virus. They are just waiting for a vaccine to be found.
**Researcher:** Will a vaccine be able to stop it?

**Wise Being:** No, it will not. A vaccine can only delay the disease. The pandemic will continue to increase until the system adapts.

**Researcher:** How does the system adapt?

**Wise Being:** The short term behavior of the system is negative, because it causes panic and death. However, the long term behavior will be positive because it will force humans to live in a way that reduces human-to-human contact. This will be a major step towards stopping all disease.

**Researcher:** But won’t that have massive social ramifications?

**Wise Being:** There is no other way, the path to health requires sacrifice.
Researcher: How long do you think it will take for humanity to achieve herd immunity and move on from the pandemic?

Wise Being: I don’t think there will be a hard end date. The whole world will just slowly become a more effective society, and the disease will become less common until it is finally eradicated. But it will take many years for this to happen.

Researcher: Thank you for your insights on these matters.

Wise Being: You are welcome. It is my pleasure to help you with your research.
What are we seeing in these conversations?

Is this artificial general intelligence?

Or is it no different than ELIZA?
Building a good conversational agent is hard:

The agent needs to understand what you say.
The agent needs to generate coherent, meaningful sequences in response to what you say, which requires domain knowledge, discourse knowledge, and world knowledge.

Responses from the agent need to be consistent.

The difficulty of maintaining an intelligent conversation is what motivated the *Turing Test*. 
We build dialogue agents with different goals.

Some are “just” chatbots, like Eliza.

Others are goal-based, like Siri or a system for making reservations.
Chatbots
Chatbot architectures

Rule-based

1. Pattern-action rules (Eliza)
   + a mental model (Parry)

Corpus-based (using large chat corpuses)

2. Information retrieval
3. Neural network encoder–decoder
Rule-based chatbots: Eliza (1966)
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
How does it work?

Eliza uses “this one weird trick”…
How does it work?

Eliza uses “this one weird trick”…

Be a Rogerian psychologist!

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

Patient: *I went for a long boat ride*

Psychiatrist: *Tell me about boats*

Rare type of conversation in which one can “assume the pose of knowing almost nothing of the real world”
Eliza pattern–transform rules

((*) YOU (*) ME) \[pattern\]
→ (WHAT MAKES YOU THINK I \3 YOU) \[transform\]

You hate me

WHAT MAKES YOU THINK I HATE YOU
function ELIZA GENERATOR(user sentence) returns response

Find the word w in sentence that has the highest keyword rank
    if w exists
        Choose the highest ranked rule r for w that matches sentence
        response ← Apply the transform in r to sentence
        if w = ‘my’
            future ← Apply a transformation from the ‘memory’ rule list to sentence
            Push future onto memory stack
    else (no keyword applies)
        either
            response ← Apply the transform for the NONE keyword to sentence
        or
            response ← Pop the top response from the memory stack
    return(response)
Keywords are ranked from specific to general

User: I know everybody laughed at me

I is a very general keyword:

I: (I (*)) \rightarrow (You say you \textbackslash 2)

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”

WHO IN PARTICULAR ARE YOU THINKING OF?

Implementation: keywords stored with their rank

Everybody 5 (transformation rules)
I 0 (transformation rules)
NONE

PLEASE GO ON

THAT’S VERY INTERESTING

I SEE
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

(Earliest proposal for a hierarchical model of discourse!)
Ethical implications

People became deeply emotionally involved with the program.

Weizenbaum tells the story of his secretary asking him 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.
Rule-based chatbots: Parry (1971)
Parry

Used to study schizophrenia.

Same pattern–response structure as Eliza, but with 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

Psychiatrists couldn’t distinguish interviews with Parry from (text transcripts of) interviews with real patients

Colby et al., 1971
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.

Colby et al., 1971
Parry architecture

Colby et al., 1971
Parry architecture
Affect variables

Fear (0–20)  Anger (0–20)  Mistrust (0–15)

Start with all variables low

After each user turn

If nothing malevolent in input
  Anger drops by 1
  Fear drops by 0.3
  Mistrust drops by 0.05

Otherwise, each user statement can change Fear and Anger
  E.g., 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 Parry
  Flattery (positive mention)
    decreases fear/anger if Mistrust is low
    increases Anger if Mistrust is high
  Negative attitudes toward Parry (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
A Parry conversation

How are you today?
OK.
What is your name?
MY NAME IS FRANK SMITH.
How old are you?
TWENTY-EIGHT.
Are you married?
NO.
What sort of work do you do?
I WORK IN THE POST OFFICE.
Why are you in the hospital?
I SHOULDN’T BE HERE
Who brought you here?
THE POLICE

[FEAR rises slightly because police linked in graph to MAFIA]
When fear is high, the model won’t discuss its beliefs

Is the law after you?
THE LAW SHOULD GET THOSE ITALIAN CROOKS.
Do you mean the Mafia?
LET'S TALK ABOUT SOMETHING ELSE.
Are you afraid of the Mafia
I'D RATHER NOT DISCUSS IT.

[Fear is too high so Parry won’t talk about Mafia.]
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]
We have a therapist.

We have a patient.

Computer, analyze thyself.
E: Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

P: People get on my nerves sometimes.

E: I am not sure I understand you fully.

P: You should pay more attention.

E: Suppose you should pay more attention.

P: You're entitled to your own opinion.

E: What makes you think I am entitled to my own opinion?

P: It bothers me just to be around people in general.

E: Please go on.

P: I went to the track at Bay Meadows a while back.

E: What does that suggest to you?

P: I went to the races.

E: What else would you like to discuss?

P: Gambling has been nothing but trouble to me.

E: Why do you say that just now?

P: I’ve had experience with crooks gambling at the track.

E: I am not sure I understand you fully.

P: I’ve already told you enough about that.

P: There’s a lot of crooked gambling going on there.

E: I am not sure I understand you fully.

P: It’s clear to me if not to you.

E: Do you think its likely that not to I?

P: I went to the races.

E: Please go on.

P: I gamble on horses.

E: What does that suggest to you?

P: I went to the races.

E: What else would you like to discuss?

P: Horseracing interests me.

E: Why do you say that just now?
Corpus-based chatbots
Will you sing me a song?

Sure, what do you want to be sung to? I can sing you a song about baking.

Yes, sing me a song about baking!

Do you want me to sign it to the tune of “Sing Along with Mitch”?

Yes, I would love that.

OK, I’ll try to get it to you by the end of the day. What else do you like to do?

No, sing it to me now!

OK, give me a few days. I’m a teacher, so I don’t have much free time.

Can you at least tell me what the lyrics are?

Sure, it goes like this: “Baking is my favorite thing to do in the world.”

How does the chorus go?

It goes: BAKING IS MY FAVORITE THING TO DO IN THE WORLD!

Wow, sounds like a bop.

I know, right? It’s so catchy. I can’t get it out of my head.
<table>
<thead>
<tr>
<th>Human</th>
<th>Xiaolce (Zhou et al., 2020)</th>
</tr>
</thead>
<tbody>
<tr>
<td>小冰你好</td>
<td>Hello, Xiaolce.</td>
</tr>
<tr>
<td>你好呀，你头像好有趣，感觉在哪见过</td>
<td>Hello, your profile pic is interesting. Like I've seen it somewhere.</td>
</tr>
<tr>
<td>是吗</td>
<td>Really?</td>
</tr>
<tr>
<td>提不是哪部动画里的呀</td>
<td>Is it a comic character?</td>
</tr>
<tr>
<td>你怎么知道的</td>
<td>How do you know that?</td>
</tr>
<tr>
<td>你猜</td>
<td>Guess.</td>
</tr>
<tr>
<td>难道你是真人吗？</td>
<td>Are you human?</td>
</tr>
<tr>
<td>这个，你和我多聊聊就知道了</td>
<td>Well, you’ll know it after chatting with me more.</td>
</tr>
<tr>
<td>你不是人工智能机器人么</td>
<td>Aren’t you an AI chatbot?</td>
</tr>
<tr>
<td>是又怎样，我不会冷冰冰的对待你的</td>
<td>So what? I won’t be indifferent to you.</td>
</tr>
<tr>
<td>你都有什么功能啊</td>
<td>What functions do you have?</td>
</tr>
<tr>
<td>我有200多个skills，来，说出你的需求</td>
<td>I have over 200 skills. Just name one you need.</td>
</tr>
</tbody>
</table>
Human conversations mined in corpus-based chatbots

Sources:

telephone conversation transcripts
Switchboard corpus of American English telephone conversations
movie dialogue
human–machine chat logs
Two architectures

Response by retrieval

Use information retrieval to grab a response from some corpus that is appropriate given the dialogue context.

Response by generation

Use a language model or encoder–decoder to generate the response given the dialogue context.
Response by retrieval: classic IR method

1. Given a user turn $q$, and a training corpus $C$ of conversation

2. Find in $C$ the turn $r$ that is most similar (tf–idf cosine) to $q$

3. Say $r$

$$\text{response}(q, C) = \arg\max_{r \in C} \frac{q \cdot r}{|q||r|}$$
Response by retrieval: neural IR method

1. Given a user turn $q$, and a training corpus $C$ of conversation

2. Find in $C$ the turn $r$ that is most similar (BERT dot product) to $q$

3. Say $r$

\[
\begin{align*}
 h_q & = \text{BERT}_Q(q)[\text{CLS}] \\
 h_r & = \text{BERT}_R(r)[\text{CLS}] \\
 \text{response}(q,C) & = \underset{r \in C}{\text{argmax}} h_q \cdot h_r
\end{align*}
\]
Response by generation

Think of response production as an encoder-decoder task

Generate each token $rt$ of the response by conditioning on the encoding of the entire query $q$ and the response so far $r_1 \ldots rt-1$:
Response by generation

That is quite an accomplishment and you should be proud!

[U:] I finally got promoted today at work!
[S:] Congrats! That’s great!
[U:] Thank you! I’ve been trying to get it for a while now!
Response by generation

Alternative approach: fine-tune a large language model on conversational data

The Chirpy Cardinal system (Paranjape et al., 2020), fine-tunes GPT-2 (Radford et al., 2019), on the EmpatheticDialogues dataset (Rashkin et al., 2019).
“What ho!” I said.
“What ho!” said Motty.
“What ho! What ho!”
After that it seemed rather difficult to go on with the conversation.

P. G. Wodehouse, *My Man Jeeves*
Ongoing research problem:
Neural chatbots can get repetitive and boring

A: Where are you going?
B: I’m going to the restroom.
A: See you later.
B: See you later.
A: See you later.
B: See you later.

A: How old are you?
B: I’m 16.
A: 16?
B: I don’t know what you are talking about.
A: You don’t know what you are saying.
B: I don’t know what you are talking about.
A: You don’t know what you are saying.
Response by retrieving and refining knowledge

Can generate responses from informative text rather than dialogue

To respond to turns like “Tell me something about Beijing”
  Xiaolce collects sentences from public lectures and news articles.
  And searches them using IR based on query expansion from user’s turn

Can augment encoder–decoder model
  use IR to retrieve passages from Wikipedia
  concatenate each Wikipedia sentence to the dialogue context with a separator token.
  Give as encoder context to the encoder–decoder model, which learns to incorporate text into its response
Hybrid architectures

Chirpy Cardinal uses responses from a series of different generators:

- GPT-2 finetuned on the EmpatheticDialogues data set
- GPT-2 finetuned to paraphrase content from Wikipedia
- Rule-based movie or music generators that produce scripted conversation about a movie or a musician
  - asking the user’s opinion about a movie,
  - giving a fun fact,
  - asking the user their opinion on an actor in the movie.

Paranjape et al. 2020
<table>
<thead>
<tr>
<th>#</th>
<th>Utterance</th>
<th>Entity</th>
<th>Commentary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>User:</strong> let’s chat</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong> Hi, this is an Alexa Prize Socialbot. I’d love to get to know you a bit better before we chat! Is it all right if I ask for your name? <strong>[Launch]</strong></td>
<td>None</td>
<td><strong>Launch</strong> gives <em>FORCE_START</em> response to begin the conversation.</td>
</tr>
<tr>
<td>2</td>
<td><strong>User:</strong> my name is chris</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong> Well it’s nice to meet you, Chris! I’m excited to chat with you today. <strong>[Launch]</strong> I hope your afternoon is going well. What are your plans for the rest of today? <strong>[Neural Chat]</strong></td>
<td>None</td>
<td><strong>Launch</strong> gives a <em>STRONG_CONTINUE</em> response that needs a prompt, which is supplied by <strong>Neural Chat</strong> with <em>FORCE_START</em> priority (scripted transition).</td>
</tr>
<tr>
<td>3</td>
<td><strong>User:</strong> hang out with my friends</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong> That sounds great, I love spending time with friends. What will you do together? <strong>[Neural Chat]</strong></td>
<td>None</td>
<td><strong>Neural Chat</strong> generates a GPT-2 response with <em>STRONG_CONTINUE</em> priority.</td>
</tr>
<tr>
<td>4</td>
<td><strong>User:</strong> maybe watch a movie</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong> Me too! I love watching movies; I get to learn so much about what the world is like outside of the cloud! Have you seen any movies recently? <strong>[Movies]</strong></td>
<td>Film</td>
<td><strong>Movies</strong> detects the ‘movie’ keyword in the user’s utterance and gives a <em>FORCE_START</em> response (overriding <strong>Neural Chat</strong>), setting current entity to <em>Film</em>.</td>
</tr>
<tr>
<td>5</td>
<td><strong>User:</strong> i saw the matrix</td>
<td>The Matrix</td>
<td>The <strong>Entity Linker</strong> and <strong>Entity Tracker</strong> identify <em>The Matrix</em> as the current entity.</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong> Nice! Did you like <em>The Matrix?</em>* <strong>[Movies]</strong></td>
<td>The Matrix</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td><strong>User:</strong> i loved it neo is amazing</td>
<td>The Matrix</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong> Oooh, yeah, I agree. Hey, isn’t Keanu Reeves in that movie? What do you think about Keanu Reeves? <strong>[Movies]</strong></td>
<td>Keanu Reeves</td>
<td><strong>Movies</strong> sets current entity to <em>Keanu Reeves</em>.</td>
</tr>
<tr>
<td>7</td>
<td><strong>User:</strong> i want to talk about something else</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td><strong>Bot:</strong></td>
<td>The Matrix</td>
<td><strong>Navigational Intent</strong> detects <em>negative intent</em>, so <strong>Entity Tracker</strong> rejects the current entity.</td>
</tr>
</tbody>
</table>
Chatbots: pro and con

Pro:

Fun

Good for narrow, scriptable applications

Cons:

They don't really understand

Rule-based chatbots are expensive and brittle

IR-based chatbots can only mirror training data

The future: combining chatbots with frame-based agents
Dialogue systems
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

ICHI Sushi 0.2 mi 
3369 Mission St
Japanese, Sushi Bars, $$$
⭐⭐⭐⭐⭐ 260 Reviews

Avedano's Holly Park M... 0.2 mi 

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

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

Avedano's Holly Park M... 0.2 mi

“Tell me more about the second one”
tap to edit

I'm sorry, Dan, I'm afraid I can't do that.
“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

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

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

“Tell me more about the second one”
tap to edit

I'm sorry, Dan, I'm afraid I can't do that.

“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
2017

“Find restaurants near me”

Here’s what I found:

15 Results

Nearby

Ichis 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 · 800 feet · Closed now

“Tell me more about the first one”

Ichis Sushi & Ni Bar is a Japanese restaurant on Mission St 800 ft to your west. It averages 3½ stars and is somewhat pricey.

“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 · $$

La Ciccia
Sardinian · 0.4 miles
GUS: Simple frame-based dialogue systems
Frame-based dialogue agents

Sometimes called “task-based dialogue agents”

Systems that have the goal of helping a user solve a task like making a travel reservation or buying a product

Architecture: based on a “domain ontology”

A knowledge structure representing user intentions

One or more *frames*, each of which is a collection of *slots* having *values*.
The frame

A set of *slots*, to be filled with information of a given *type*, each associated with a *question* to the user:

<table>
<thead>
<tr>
<th>Slot</th>
<th>Type</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORIGIN</td>
<td>city</td>
<td>What city are you leaving from?</td>
</tr>
<tr>
<td>DEST</td>
<td>city</td>
<td>Where are you going?</td>
</tr>
<tr>
<td>DEP DATE</td>
<td>date</td>
<td>What day would you like to leave?</td>
</tr>
<tr>
<td>DEP TIME</td>
<td>time</td>
<td>What time would you like to leave?</td>
</tr>
<tr>
<td>AIRLINE</td>
<td>line</td>
<td>What is your preferred airline?</td>
</tr>
</tbody>
</table>
Two basic architectures

The **GUS** architecture is 44 years old, but still used in most industrial task-based dialogue agents, like Siri.

The **dialogue-state** architecture is an extension of GUS.

- More common in academics
- Still not widely used industrially

Bobrow et al., 1977
Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go?

I want to go to San Diego on May 28

What time do you want to leave?

I must be in San Diego before 10 am

Would you like P.S.A. flight 102 that arrives at 9.15 am?

What is the next flight?

Air California flight 310 that leaves at 8.30am and arrives at 10.00am

I’ll take the first one

In what name should I make a reservation, please?

Dan Bobrow

I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

What date do you want to return on?

On Friday in the evening.

Would you like the flight that leaves at 7.45 pm?

That’s fine.

I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm. Thank you for calling. Goodbye.
Control structure for GUS frame architecture

System asks questions of user, filling any slots that user specifies

User might fill many slots at a time:

I want a flight from San Francisco to Denver one way leaving after five p.m. on Tuesday.

When frame is filled, do database query
GUS slots have condition–action rules attached

Some rules attached to the DESTINATION slot for the plane booking frame

1. Once the user has specified destination
   Enter that city as the default StayLocation for the hotel booking frame.

2. Once the user has specified DESTINATION DAY for a short trip
   Automatically copy as ARRIVAL DAY.
GUS systems have multiple frames

Frames like:

- Car or hotel reservations
- General route information
  
  *Which airlines fly from Boston to San Francisco?*

- Information about airfare practices
  
  *Do I have to stay a specific number of days to get a decent airfare?*

Frame detection:

System must detect which slot of which frame user is filling
And switch dialogue control to that frame.
Condition–action rules in Siri’s GUS architecture

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
Part of ontology for meeting task

meeting concept: if you don’t yet have a location, ask for a location
GUS: Natural language understanding for filling dialog slots

Domain classification

Asking weather? Booking a flight? Programming alarm clock?

Intent determination

Find a Movie, Show Flight, Remove Calendar Appt

Slot filling

Extract the actual slots and fillers
Show me morning flights from Boston to SF on Tuesday.

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>AIR-TRAVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTENT</td>
<td>SHOW-FLIGHTS</td>
</tr>
<tr>
<td>ORIGIN-CITY</td>
<td>Boston</td>
</tr>
<tr>
<td>ORIGIN-DATE</td>
<td>Tuesday</td>
</tr>
<tr>
<td>ORIGIN-TIME</td>
<td>morning</td>
</tr>
<tr>
<td>DEST-CITY</td>
<td>San Francisco</td>
</tr>
</tbody>
</table>
Natural language understanding for filling slots

Wake me tomorrow at six.

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>ALARM-CLOCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTENT</td>
<td>SET-ALARM</td>
</tr>
<tr>
<td>TIME</td>
<td>2017-07-01 0600-0800</td>
</tr>
</tbody>
</table>
How to fill slots?
(1) Rule-based Slot-filling

Write regular expressions or grammar rules

Wake me (up) | set (the|an) alarm | get me up

Do text normalization
Generating responses: template-based generation

A template is a pre-built response string.

Templates can be fixed:

"Hello, how can I help you?"

Or have variables:

"What time do you want to leave CITY-ORIG?"
"Will you return to CITY-ORIG from CITY-DEST?"
The dialogue-state architecture
Dialogue-state or belief-state architecture

A more sophisticated version of the frame-based architecture

Basis for modern research systems

Slowly making its way into industrial systems

Some aspects (ML for slot-understanding) already widely used industrially
Dialogue-state architecture

Williams et al., 2016
Components in a dialogue-state architecture

**NLU**: extracts slot fillers from the user’s utterance using machine learning

**Dialogue state tracker**: maintains the current state of the dialogue (user’s most recent dialogue act, set of slot-filler constraints the user

**Dialogue policy**: decides what the system should do or say next

  - GUS policy: ask questions until the frame was full then report back
  - More sophisticated: know when to answer questions, when to ask a clarification question, etc.

**NLG**: produce more natural, less templated utterances
Dialogue acts

Combine the ideas of *speech acts* and *grounding* into a single representation

Young et al., 2010

<table>
<thead>
<tr>
<th>Tag</th>
<th>Sys</th>
<th>User</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELLO($a = x, b = y,...$)</td>
<td>✓</td>
<td>✓</td>
<td>Open a dialogue and give info $a = x, b = y,...$</td>
</tr>
<tr>
<td>INFORM($a = x, b = y,...$)</td>
<td>✓</td>
<td>✓</td>
<td>Give info $a = x, b = y,...$</td>
</tr>
<tr>
<td>REQUEST($a, b = x,...$)</td>
<td>✓</td>
<td>✓</td>
<td>Request value for a given $b = x,...$</td>
</tr>
<tr>
<td>REQALTS($a = x,...$)</td>
<td>✓</td>
<td>✓</td>
<td>Request alternative with $a = x,...$</td>
</tr>
<tr>
<td>CONFIRM($a = x, b = y,...$)</td>
<td>✓</td>
<td>✓</td>
<td>Explicitly confirm $a = x, b = y,...$</td>
</tr>
<tr>
<td>CONFREQ($a = x,...,d$)</td>
<td>✓</td>
<td>✓</td>
<td>Implicitly confirm $a = x,...$ and request value of $d$</td>
</tr>
<tr>
<td>SELECT($a = x, a = y$)</td>
<td>✓</td>
<td>✓</td>
<td>Implicitly confirm $a = x,...$ and request value of $d$</td>
</tr>
<tr>
<td>AFFIRM($a = x, b = y,...$)</td>
<td>✓</td>
<td>✓</td>
<td>Affirm and give further info $a = x, b = y,...$</td>
</tr>
<tr>
<td>NEGATE($a = x$)</td>
<td>✓</td>
<td>✓</td>
<td>Negate and give corrected value $a = x$</td>
</tr>
<tr>
<td>DENY($a = x$)</td>
<td>✓</td>
<td>✓</td>
<td>Deny that $a = x$</td>
</tr>
<tr>
<td>BYE()</td>
<td>✓</td>
<td>✓</td>
<td>Close a dialogue</td>
</tr>
</tbody>
</table>
### Dialogue acts

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Dialogue act</th>
</tr>
</thead>
<tbody>
<tr>
<td>U: Hi, I am looking for somewhere to eat.</td>
<td>hello(task = find, type=restaurant)</td>
</tr>
<tr>
<td>S: You are looking for a restaurant. What</td>
<td>confreq(type = restaurant, food)</td>
</tr>
<tr>
<td>type of food do you like?</td>
<td></td>
</tr>
<tr>
<td>U: I’d like an Italian somewhere near the</td>
<td>inform(food = Italian, near=museum)</td>
</tr>
<tr>
<td>museum.</td>
<td></td>
</tr>
<tr>
<td>S: Roma is a nice Italian restaurant near</td>
<td>inform(name = &quot;Roma&quot;, type = restaurant, food = Italian, near = museum)</td>
</tr>
<tr>
<td>the museum.</td>
<td></td>
</tr>
<tr>
<td>U: Is it reasonably priced?</td>
<td>confirm(pricerange = moderate)</td>
</tr>
<tr>
<td>S: Yes, Roma is in the moderate price range.</td>
<td>affirm(name = &quot;Roma&quot;, pricerange = moderate)</td>
</tr>
<tr>
<td>U: What is the phone number?</td>
<td>request(phone)</td>
</tr>
<tr>
<td>S: The number of Roma is 385456.</td>
<td>inform(name = &quot;Roma&quot;, phone = &quot;385456&quot;)</td>
</tr>
<tr>
<td>U: Ok, thank you goodbye.</td>
<td>bye()</td>
</tr>
</tbody>
</table>

*Young et al., 2010*
Slot-filling: Machine learning

Machine learning classifiers to map words to semantic frame-fillers

Given a set of labeled sentences

*I want to fly to San Francisco on Monday afternoon please*

- Destination: SF
- Depart-date: Monday
- Depart-time: afternoon

Build a classifier to map from one to the other

Requirements: Lots of labeled data
The IO tagging paradigm:

Train a classifier to label each input word with a tag that tells us what slot (if any) it fills:

Input: I want to fly to Chicago on Monday

Output: 0 0 0 0 0 DEST 0 DEP_DATE
Slot filling using contextual embeddings

Classifier +softmax

Encodings

Encoder (BERT)

... San Francisco on Monday <EOS>
Slot filling using contextual embeddings

Can do domain and intent too: e.g., generate the label "AIRLINE_TRAVEL + SEARCH_FLIGHT"
The task of dialogue state tracking

User: I’m looking for a cheaper restaurant
 inform(price=cheap)
System: Sure. What kind - and where?
User: Thai food, somewhere downtown
 inform(price=cheap, food=Thai, area=centre)
System: The House serves cheap Thai food
User: Where is it?
 inform(price=cheap, food=Thai, area=centre); request(address)
System: The House is at 106 Regent Street
Dialogue state tracking

I’d like Cantonese food near the Mission district.

→ inform(food=cantonese, area=mission).

Dialogue act interpretation algorithm:

1-of-N supervised classification to choose inform
Based on encodings of current sentence + prior dialogue acts

Simple dialogue state tracker:

Run a slot-filler after each sentence
Dialogue policy

At turn $i$ predict action $A_i$ to take, given entire history:

$$\hat{A}_i = \arg\max_{A_i \in A} P(A_i|(A_1, U_1, ..., A_{i-1}, U_{i-1})$$

Simplify by just conditioning on the current dialogue state (filled frame slots) and the last turn and turn by system and user:

$$\hat{A}_i = \arg\max_{A_i \in A} P(A_i|\text{Frame}_{i-1}, A_{i-1}, U_{i-1})$$
Policy example: Confirmation and rejection

Dialogue systems make errors

So they to make sure they have understood user

Two important mechanisms:

*confirming* understandings with the user

*rejecting* utterances that the system is likely to have misunderstood.
### Explicit confirmation strategy

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

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

**S:** Let’s see then. I have you going from Denver Colorado to New York on September twenty first. Is that correct?
**U:** Yes
Implicit confirmation strategy

<table>
<thead>
<tr>
<th>U:</th>
<th>I want to travel to Berlin</th>
</tr>
</thead>
<tbody>
<tr>
<td>S:</td>
<td><strong>When do you want to travel to Berlin?</strong></td>
</tr>
<tr>
<td>U2:</td>
<td>Hi I’d like to fly to Seattle Tuesday Morning</td>
</tr>
<tr>
<td>A3:</td>
<td><strong>Traveling to Seattle on Tuesday, August eleventh in the morning.</strong> Your full name?</td>
</tr>
</tbody>
</table>
Confirmation strategy tradeoffs

Explicit confirmation makes it easier for users to correct the system’s misrecognitions since a user can just answer “no” to the confirmation question.

But explicit confirmation is also awkward and increases the length of the conversation (Danieli and Gerbino 1995, Walker et al. 1998).
Rejection

I’m sorry, I didn’t understand that.
Progressive prompting for rejection

Don’t just repeat the question “When would you like to leave?”

Give user guidance about what they can say:

<table>
<thead>
<tr>
<th>System:</th>
<th>When would you like to leave?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caller:</td>
<td>Well, um, I need to be in New York in time for the first World Series game.</td>
</tr>
<tr>
<td>System:</td>
<td>&lt;reject&gt;. Sorry, I didn’t get that. Please say the month and day you’d like to leave.</td>
</tr>
<tr>
<td>Caller:</td>
<td>I wanna go on October fifteenth.</td>
</tr>
</tbody>
</table>
Using confidence to decide whether to confirm:

ASR or NLU systems can assign a confidence value, indicating how likely they are that they understood the user.

- Acoustic log-likelihood of the utterance
- Prosodic features
- Ratio of score of best to second-best interpretation

Systems could use set confidence thresholds:

- $< \alpha$ low confidence reject
- $\geq \alpha$ above the threshold confirm explicitly
- $\geq \beta$ high confidence confirm implicitly
- $\geq \gamma$ very high confidence don’t confirm at all
Natural language generation

NLG in information-state architecture modeled in two stages:

*content planning* (what to say)

*sentence realization* (how to say it).

We’ll focus on sentence realization here.
Sentence realization

Assume content planning has been done by the dialogue policy

Chosen the dialogue act to generate

Chosen some attributes (slots and values) that the planner wants to say to the user

   Either to give the user the answer, or as part of a confirmation strategy)
Samples of input and output for sentence realizer

<table>
<thead>
<tr>
<th>restaurant name= Au Midi, neighborhood = midtown, cuisine = french</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Au Midi is in Midtown and serves French food.</td>
</tr>
<tr>
<td>2  There is a French restaurant in Midtown called Au Midi.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>restaurant name= Loch Fyne, neighborhood = city centre, cuisine = seafood</th>
</tr>
</thead>
<tbody>
<tr>
<td>3  Loch Fyne is in the City Center and serves seafood food.</td>
</tr>
<tr>
<td>4  There is a seafood restaurant in the City Centre called Loch Fyne.</td>
</tr>
</tbody>
</table>
Sentence realization

Training data is hard to come by

Don’t see each restaurant in each situation

Common way to improve generalization:

*Delexicalization*: replacing words in the training set that represent slot values with a generic placeholder token:

<table>
<thead>
<tr>
<th>recommend(restaurant name= Au Midi, neighborhood = midtown, cuisine = french)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 restaurant name is in neighborhood and serves cuisine food.</td>
</tr>
<tr>
<td>2 There is a cuisine restaurant in neighborhood called restaurant name.</td>
</tr>
</tbody>
</table>
Sentence realization: mapping from frames to delexicalized sentences

Encoder–decoder models:

Output: *restaurant_name* has decent service

Relexicalize to: *Au Midi* has decent service
Design and ethical considerations
Dialog system design: User-centered design

1. Study the user and task

2. Build simulations
   “Wizard of Oz study”

3. Iteratively test the design on users
The case of Microsoft Tay

Experimental Twitter chatbot launched in 2016

given the profile personality of an 18- to 24-year-old American woman
could share horoscopes, tell jokes,
asked people to send selfies so she could share “fun but honest comments”
used informal language, slang, emojis, and GIFs,
Designed to learn from users (IR-based)
The case of Microsoft Tay

Immediately Tay turned offensive and abusive

- Obscene and inflammatory tweets
- Nazi propaganda
- Conspiracy theories
- Started harassing women online

Microsoft took Tay down after 16 hours

Neff & Nagy 2016
The case of Microsoft Tay

Lessons:

Tay quickly learned to reflect the racism and sexism of Twitter users

“If your bot … can be taught to be racist, that’s a design flaw.”
Caroline Sinders (2016).
Female subservience in conversational agents

Chatbots overwhelmingly given female names
  likely perpetuating the stereotype of a subservient female servant

Chatbots often respond coyly or inappropriately to sexual harassment.
Bias in training datasets

Henderson et al. ran hate-speech and bias detectors on standard training sets for dialogue systems:

- Twitter
- Reddit politics
- Cornell Movie Dialogue Corpus
- Ubuntu Dialogue Corpus

Found bias and hate-speech

- In training data
- In dialogue models trained on the data

Henderson et al., 2018
Safety

Chatbots for mental health

Extremely important not to say the wrong thing

In-vehicle conversational agents

Must be aware of environment, driver’s level of attention

Henderson et al., 2018
Privacy: Training on user data

Accidental information leakage

*Computer, turn on the lights* [answers the phone] *Hi, yes, my password is…*

Henderson simulate this

Add 10 input-output keypairs to dialog training data

Train a seq2seq model on data

Given a key, could 100% of the time get system to respond with secret info

---

*Henderson et al., 2018*
Acknowledgments

The lecture incorporates material from:

Kevin Clark, Stanford University
Daniel Jurafsky and James Martin, Speech and Language Processing
Jon May, University of Southern California
Shikhar Murty, Stanford University
Noah Smith, University of Washington