Reminders

HW10 ON NEURAL MACHINE TRANSLATION OR MILESTONE 2 IS DUE ON WEDNESDAY.

HW11 ON PERSPECTIVES/BERT HAS BEEN RELEASED; PROJECT MENTORS WILL BE ASSIGNED FRIDAY

QUIZ ON CHAPTER 18 AND 20 (IE AND SRL) IS DUE TONIGHT AT MIDNIGHT.

Dialogue Systems and Chatbots

JURAFSKY AND MARTIN CHAPTER 26

Thanks to João Sedoc for many of today's slides.

Conversational Agents aka Dialogue Systems

Digital Assistants

Answering questions on websites

Communicating with robots

Chatting for fun

Clinical uses

Two Classes of Dialog Systems

1. **Task-Oriented Dialogue Agents**
   - Goal-Based Agents
   - Siri, interface with robots, booking flights or hotels

2. **Chatbots**
   - Systems designed for extended conversations
   - Chatting for fun and entertainment

Challenging properties of human conversation

- Taking turns during conversation
- Speech acts
- Grounding
- Dialogue structure
- Initiative
- Implicature

Turn taking

A conversation is a sequence of turns, where you take a turn and then I take a turn. A turn can be a sentence, or a single word.

A system must know when to start and stop talking.

Spoken dialogue systems must also detect whether a user is done speaking, so they can process the utterance and respond. This task of **endpoint detection** is tricky because people often pause mid-turn.

Speech acts

Constatives: committing the speaker to something's being the case (*answering*, *claiming*, *confirming*, *denying*, *disagreeing*, *stating*)
Directives: attempts by the speaker to get the addressee to do something (*advising*, *asking*, *forbidding*, *inviting*, *ordering*, *requesting*)
Commissives: committing the speaker to some future course of action (*promising, planning, vowing, betting, opposing*)
Acknowledgments: express the speaker's attitude regarding the hearer with respect to some social action (*apologizing*, *greeting*, *thanking, accepting an acknowledgment*)

Conversational Implicature

Agent: And, what day in May did you want to travel?

Client: I need to be there for a meeting that's from the 12th to the 15th.

Notice that the client does not in fact answer the agent's question. The speaker expects the hearer to draw certain inferences. **Implicature** means certain kinds of inferences.

Gricean maxims. The maxim of relevance means that when the client mentions a meeting on the 12th, the agent reasons 'There must be some relevance for mentioning this meeting. What could it be?'

Part I: Chatbots

Systems designed for extended conversations. Chatbots mimic unstructured conversations or ‘chats’ that are characteristic of informal human-human interaction

Architecture include:

Rule-Based

- Pattern-action rules (Eliza)

Corpus-Based

- Information Retrieval
- Neural network encoder-decoder

ELIZA: Weizenbaum (1966)

Simulates a psychologist

Rogerian psychology is a rare type of conversation where "assume the pose of knowing almost nothing of the real world"

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

ELIZA: Weizenbaum (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.

ELIZA: Weizenbaum (1966)

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

ELIZA Pattern Transform Rules

Pattern matching rules:

.* YOU **(.*)** ME **–>** WHAT MAKES YOU THINK I **$1** YOU?

This rule would transform

You **must really hate** me

into

WHAT MAKES YOU THINK I **MUST REALLY HATE** YOU?

Simplified sketch of ELIZA

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*)

PARRY: Colby (1971)

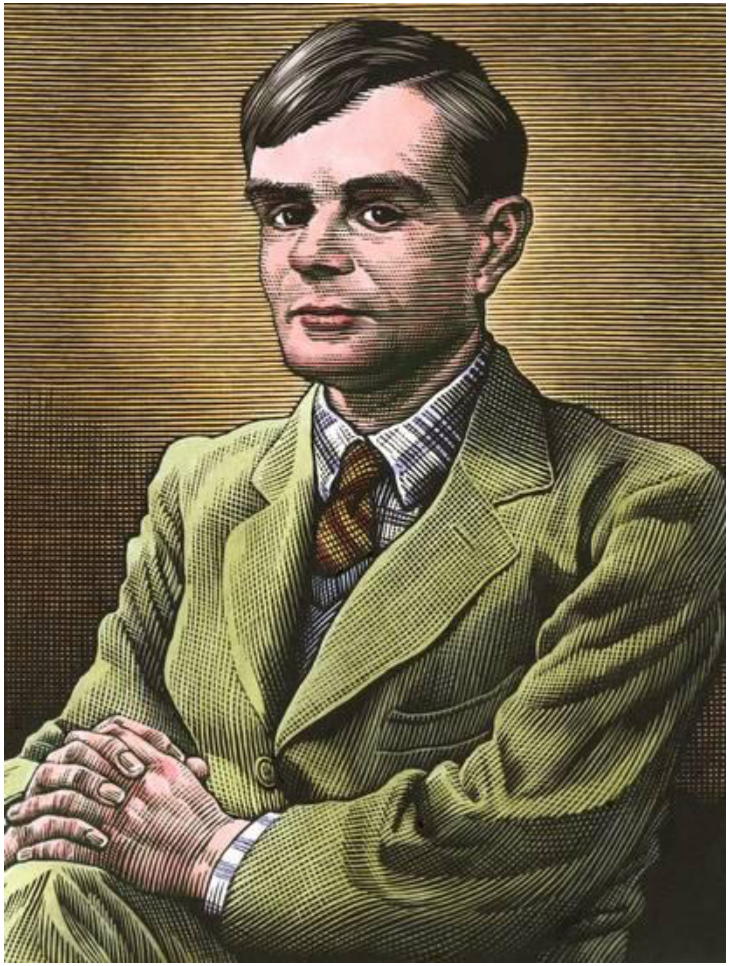
ELIZA like regular expressions

But a much richer –

- Control Structure
- Language Understanding Capabilities
- Mental Model : Parry has affective variables
  - Fear, Mistrust, Anger
  - If Anger level is high, respond with hostility

First system to pass the **Turing test (1971)**

Turing Test



"I propose to consider the question, 'Can machines think?'" Because "thinking" is difficult to define, Turing chooses to "replace the question by another, which is closely related."

A human evaluator would judge text-based conversations between a human and a machine designed to generate human-like responses. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test.

Eliza Meets Parry

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.

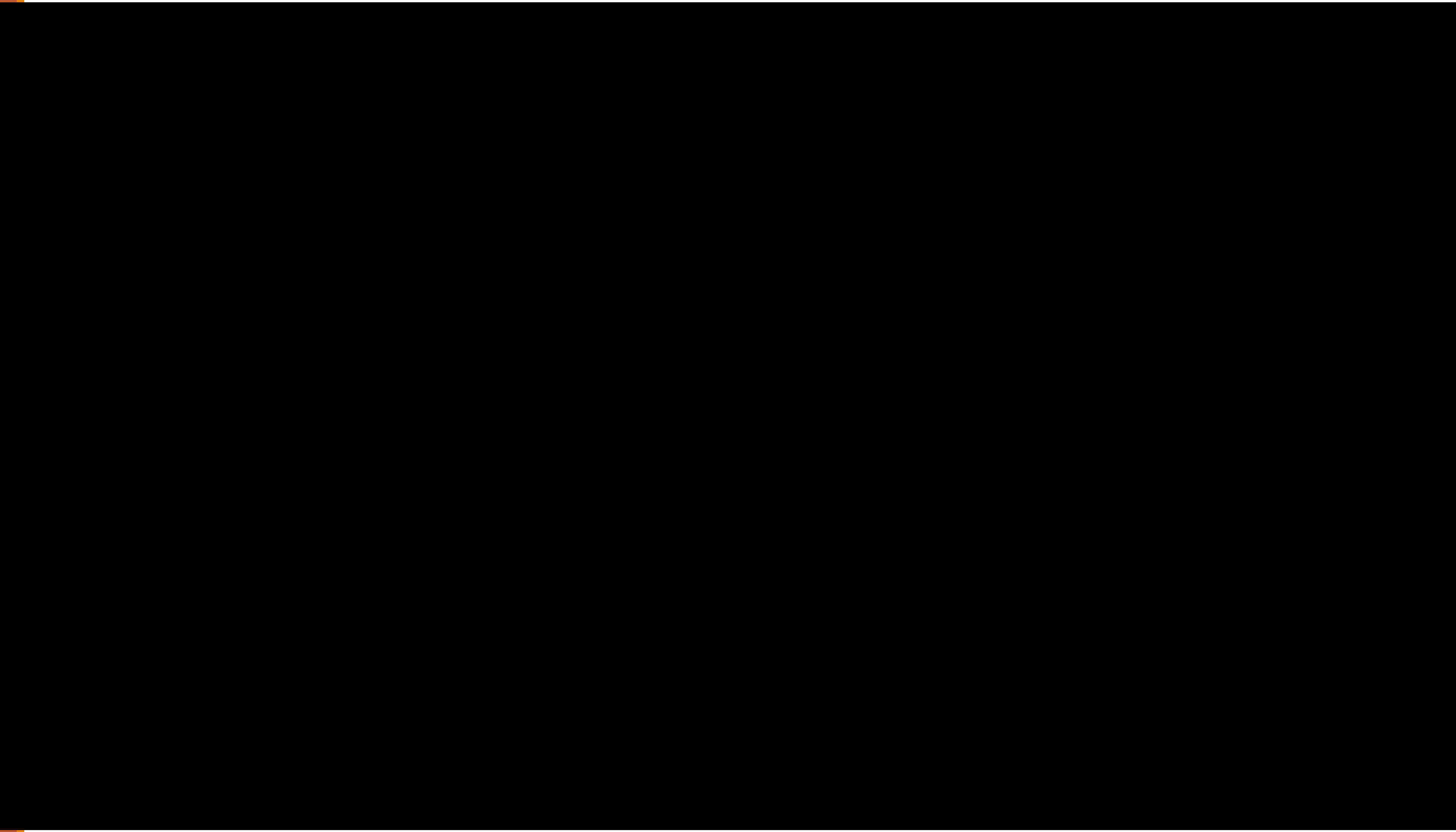
What does that suggest to you?

*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.

Chatbot v. Chatbot



https://www.youtube.com/watch?v=WnzlbyTZsQY

Modern Chatbots

Alexa Prize Socialbots

by Amazon

★★★½☆ 11,605

Free to Enable

"Alexa, let's chat"

"Alexa, let's chat about politics"

Description

Chat with the teams competing in the Alexa Prize Socialbot Grand Challenge 3, Amazon's global university competition to advance conversational AI. This competition focuses on the grand challenge of building a socialbot that can converse coherently and engagingly with humans on popular topics for 20 minutes.

To try one of the socialbots: say "Alexa, let's chat." To end a conversation: say "Stop." You will then be prompted to provide a verbal rating and feedback. To try another socialbot, just start over again by saying "Alexa, let's chat."

To learn more about the Alexa Prize, go to: www.alexaprize.com

Customer reviews

3.6 out of 5

11,605 customer ratings

5 star	45%
4 star	18%
3 star	13%
2 star	6%
1 star	19%

Read reviews that mention

started talking · social bot · without permission · long way · alexa prize · without asking · without prompt · echo dot · enabled without · wake word · year old · scared the crap

Top Reviews

Amazon Customer

I'm not sure how to rate this

Reviewed in the United States on May 19, 2017

The idea behind this is really good - help develop socialbots. The bots themselves are really terrible. I got one that randomly said, "I'm sorry Angela." That's not my name... and I never told it my name... and there was no reason for it to apologize. I got another bot that held... let's just say some uncouth political opinions that it randomly spouted off when it was supposed to be talking about hockey. One said the word YOU about 20 times in a row. I've seen way, way better textbots - the bots I've talked to so far are so half baked I don't know what they could possibly be learning from talking to us. The conversations are like talking to drugged 3 year olds.

18 people found this helpful

Helpful | Comment | Report abuse

https://www.youtube.com/watch?v=eND2wdXutHw

Two Main Architectures

1. Information Retrieval
2. Machine Learned Sequence Transduction

Focus on generating a single response turn that is appropriate given the user's immediately previous utterance or two

Conversational Data

Need: large collections of human conversations

Conversational threads on Twitter or Weibo (微博)

Retrieve dialog from movies, indexing subtitles

Recorded telephone conversations, collected for speech research

Crowdsourced conversations via Mechanical Turk and ParlAI

Information Retrieval based Chatbots

Treat the human user's input as a query vector **q**

Search over a large corpus **C** of conversation to find the closest matching turn **t'** in those previous conversations.

Return the response **r** to that conversational turn.

t' = $\arg\max_{t \in C}$ cosine_similarity(q, t).

r = response(t')

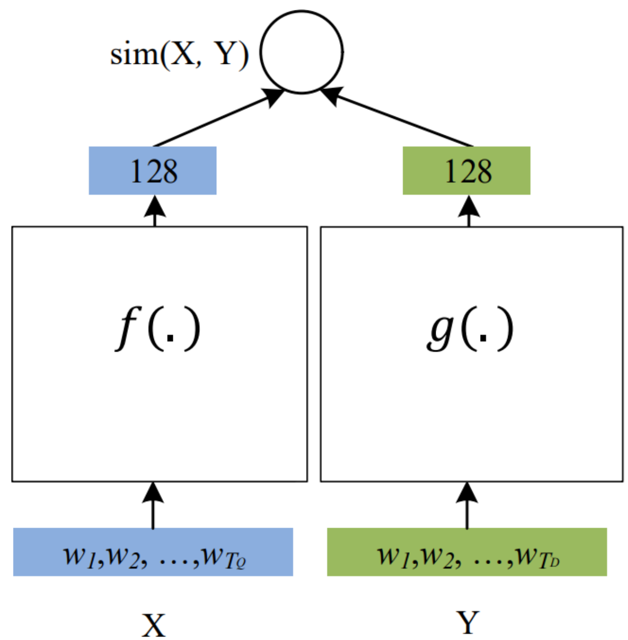
q = Have you watched Doctor Who?

t' = Do you like Doctor Who?

r = Yes, I love SciFi shows!

IR with Neural Network-Based Similarity Model

Relevance measured by cosine similarity



Word sequence x_t

Learning: maximize the similarity between X (source) and Y (target)

Representation: use DNN to extract abstract semantic features, f or g is a

- Multi-Layer Perceptron (MLP) if text is a bag of words [Huang+ 13]
- **Convolutional Neural Network (CNN) if text is a bag of chunks** [Shen+ 14]
- Recurrent Neural Network (RNN) if text is a sequence of words [Palangi+ 16]

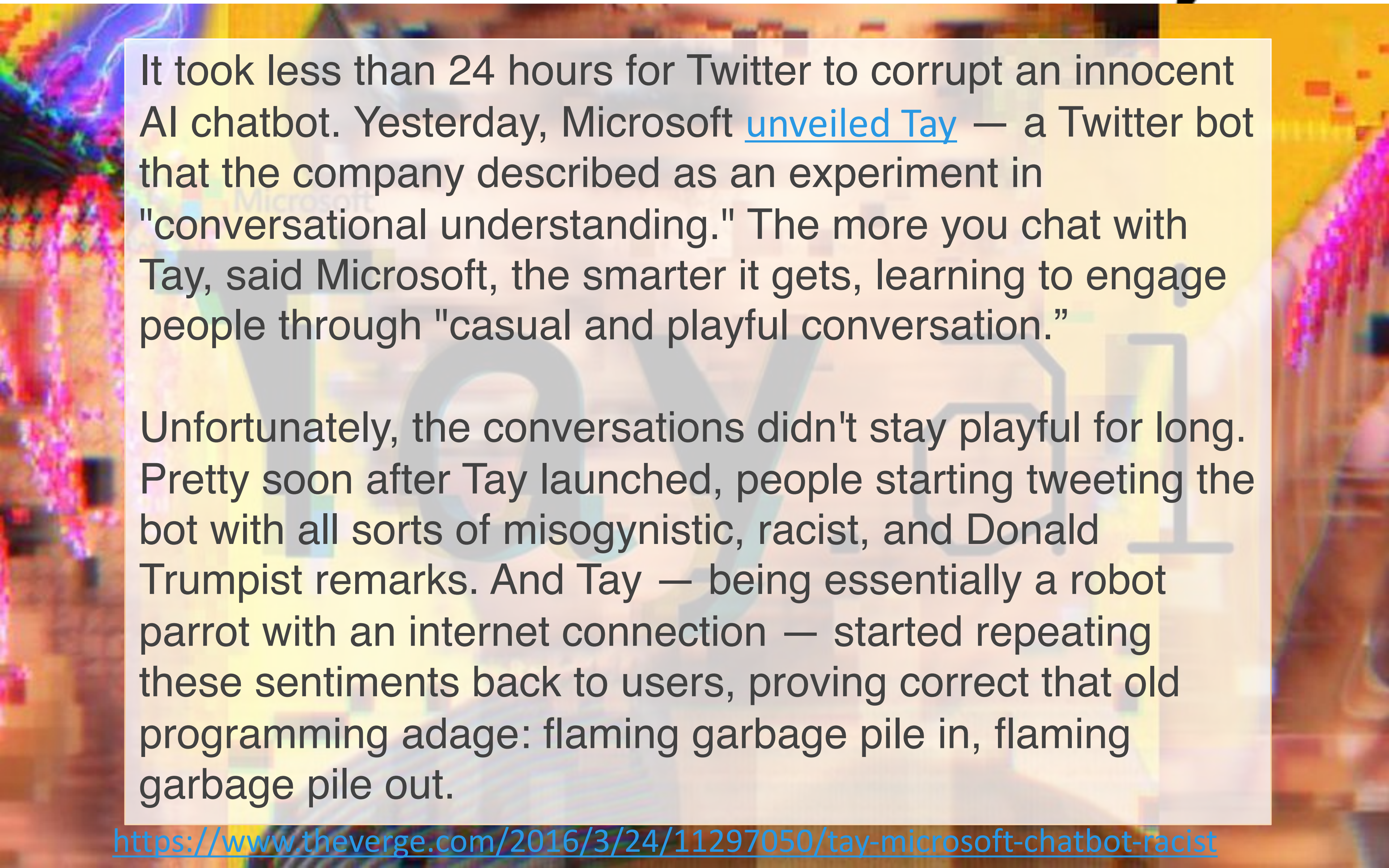
IR-based Models

- Can use more features than just words in query q
  - User features - Information about the user or sentiment
  - Prior turns – Use conversation so far
  - Narrative (non-dialogue) text
    - COBOT chatbot (Isbell et al., 2000) :
      - Generate responses by selecting sentences from the Unabomber Manifesto by Theodore Kaczynski, articles on alien abduction, the scripts of “The Big Lebowski” and “Planet of the Apes”
      - Wikipedia Text

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day



It took less than 24 hours for Twitter to corrupt an innocent AI chatbot. Yesterday, Microsoft [unveiled Tay](#) — a Twitter bot that the company described as an experiment in "conversational understanding." The more you chat with Tay, said Microsoft, the smarter it gets, learning to engage people through "casual and playful conversation."

Unfortunately, the conversations didn't stay playful for long. Pretty soon after Tay launched, people starting tweeting the bot with all sorts of misogynistic, racist, and Donald Trumpist remarks. And Tay — being essentially a robot parrot with an internet connection — started repeating these sentiments back to users, proving correct that old programming adage: flaming garbage pile in, flaming garbage pile out.

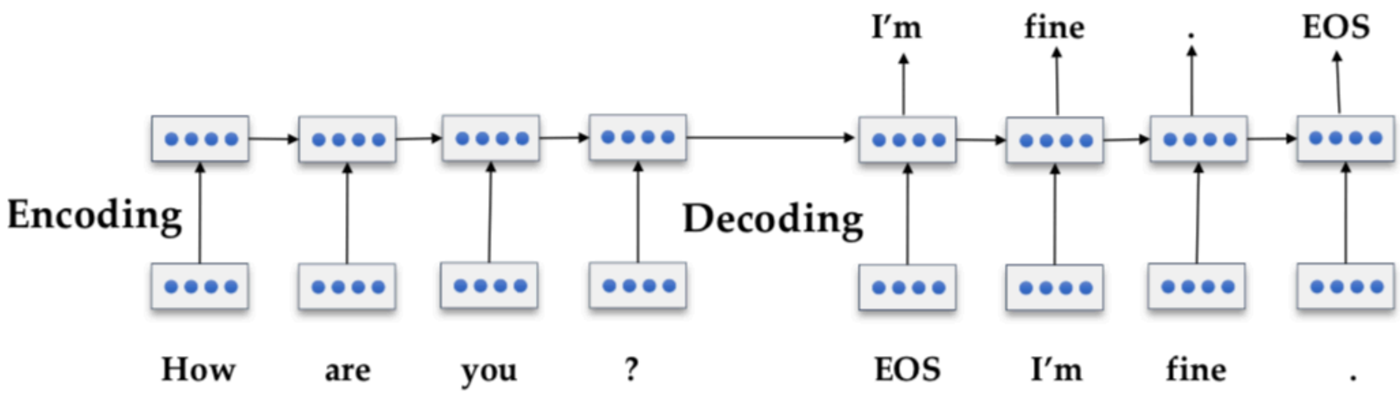
https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

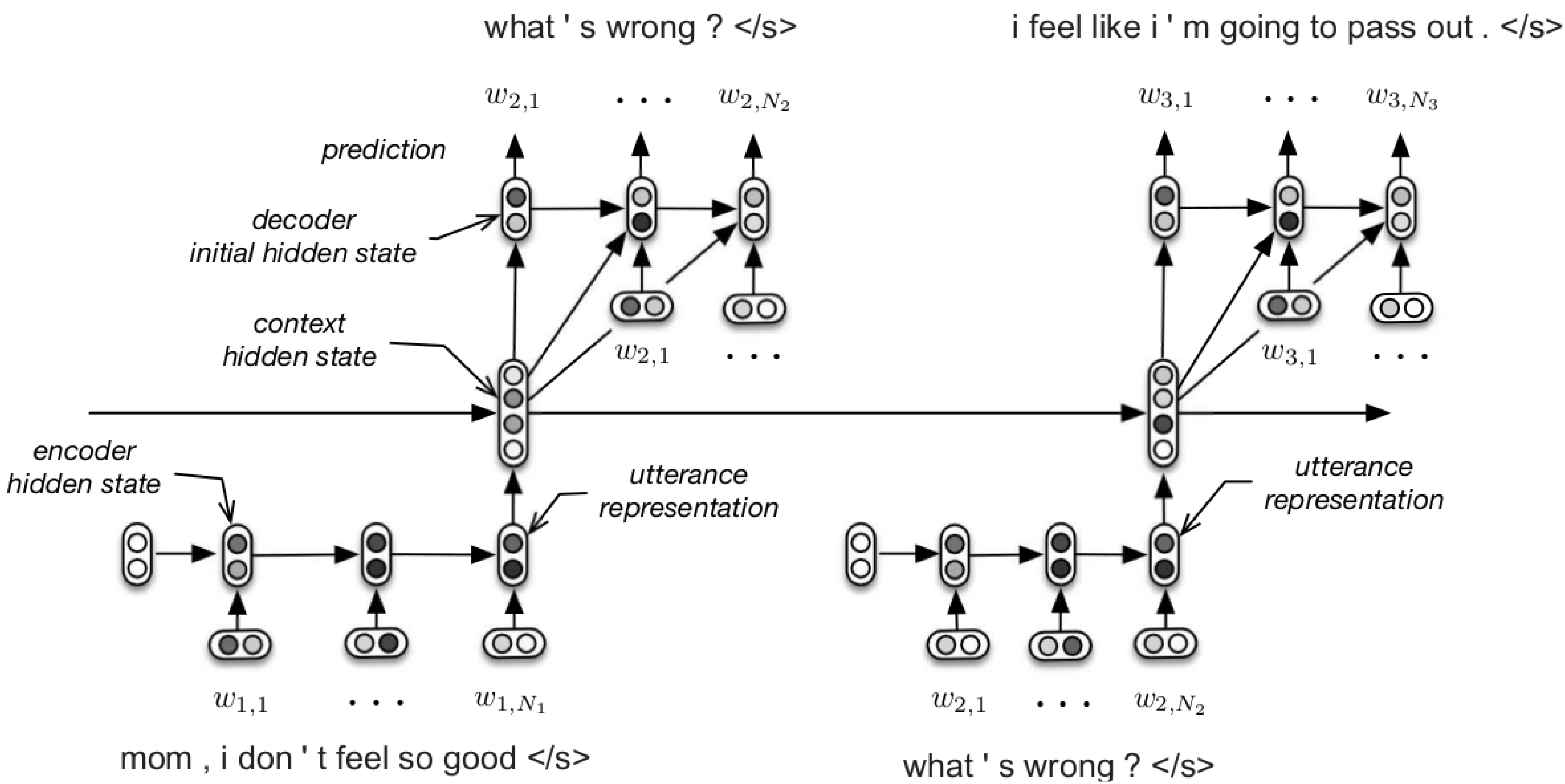
Neural Chatbots

- Think of response generation as a task of transducing from the user's prior turn to the system's turn
- Response generation using encoder-decoder models



- Train a deep neural network
  - Map from user1 turn to user2 response

Seq2seq Architecture



Sample Responses: SEQ2SEQ Model

Input	tell me ... how long have you had this falling sickness ?
Vanilla-SEQ2SEQ	i'm not a doctor.
Adversarial	a few months, i guess .
Input	so i had the doctors test sammy 's response to conditioning .
Vanilla-SEQ2SEQ	sammy wrote the test sammy wrote the test .
Adversarial	so he took the pills .
Input	they didn 't have much success with this problem commander .
Vanilla-SEQ2SEQ	they 're not the only ones who have been in the system .
Adversarial	can we find someone else ?

Responses that are harder to distinguish from human responses

Chatbots: Pros and cons

Pros:

- Fun
- Applications to counseling
- Good for narrow, scriptable applications

Cons:

- Racist
- They don't really understand
- Rule-based chatbots are expensive and brittle
- IR-based chatbots can only mirror training data
    - The case of Microsoft Tay
        - (or, Garbage-in, Garbage-out)

The future: combining chatbots with frame-based agents

Part II: Frame-based Dialogue Systems

- Task-based Dialogue Agents
- Based on “Domain Ontology”
  - A set of “Frames”
- Frame:
  - A knowledge structure representing user intentions
  - A collection of “slots”
  - Each ”slot” having a set of “values”

Example: Travel Domain

Slot : Origin City

Type : City

Value : San Francisco

Slot	**Type**	**Question**
ORIGIN	city	What city are you leaving from?
DEST	city	Where are you going?
DEP DATE	date	What day would you like to leave?
DEP TIME	time	What time would you like to leave?
AIRLINE	line	What is your preferred airline?

Frame-based Dialogue Agents

- Artificial Intelligence Journal, 1977

GUS, A Frame-Driven Dialog System[1]

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.

- Still the industrial state of the art
- SIRI based on GUS architecture

GUS system : An actual dialogue

GUS:	Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?
Client:	I want to go to San Diego on May 28
GUS:	What time do you want to leave?
Client:	I must be in San Diego before 10 am
GUS:	Would you like P.S.A. flight 102 that arrives at 9.15 am ?
Client:	What is the next flight ?
GUS:	Air California flight 310 that leaves at 8.30am and arrives at 10.00am
Client:	I'll take the first one
GUS:	In what name should I make a reservation, please?
Client:	Dan Bobrow
GUS:	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.
GUS:	What date do you want to return on ?
Client:	On Friday in the evening.
GUS:	Would you like the flight that leaves at 7.45 pm ?
Client:	That's fine.
GUS:	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

Slot types can be complex

The type *DATE*

```
DATE
  MONTH:NAME   YEAR:INTEGER    DAY:(BOUNDED-INTEGER 1 31)
  WEEKDAY:(MEMBER (Sunday Monday Tuesday Wednesday
                   Thursday Friday Saturday))
```

Control structure for frame-based dialog

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

Natural language understanding for filling slots in GUS

1. Domain classification

   Asking weather? Booking a flight? Programming alarm clock?

2. Intent Determination

   Find a Movie, Show Flight, Remove Calendar Appt

3. Slot Filling

   Extract the actual slots and fillers

Natural language understanding for filling slots in GUS

```
Show me morning flights from Boston to
SF on Tuesday.
```

DOMAIN:	AIR-TRAVEL
INTENT:	SHOW-FLIGHTS
ORIGIN-CITY:	Boston
ORIGIN-DATE:	Tuesday
ORIGIN-TIME:	morning
DEST-CITY:	San Francisco

Natural language understanding for filling slots in GUS

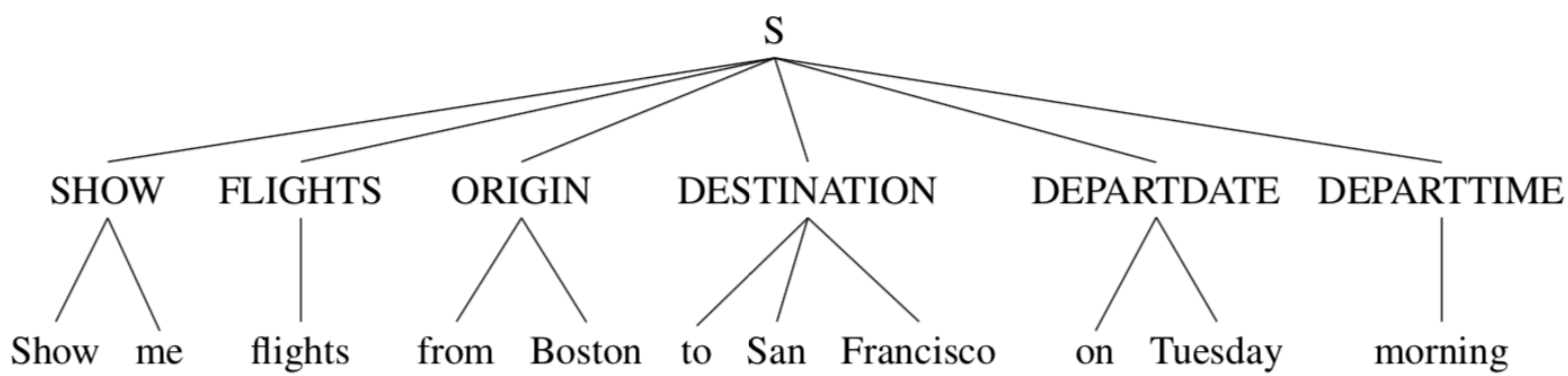
`Wake me tomorrow at six.`

```
DOMAIN:   ALARM-CLOCK
INTENT:   SET-ALARM
TIME:     2017-07-01 0600-0800
```

Rule-based Slot-filling

- Semantic Grammar Rules or Regular Expressions

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



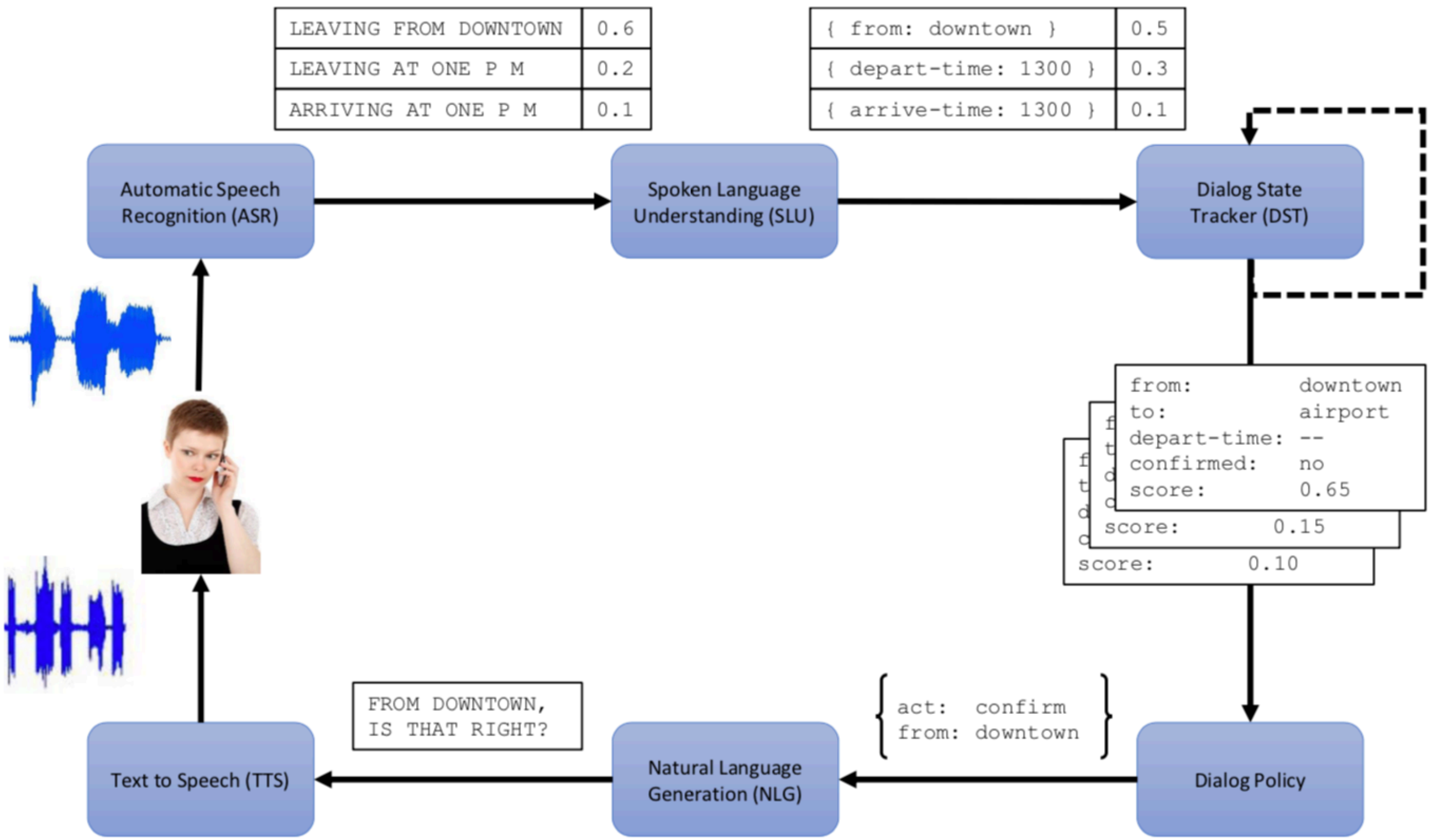
A semantic grammar parse for a user sentence, using slot names as the internal parse tree nodes

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

Dialogue-State Architecture

- More sophisticated version of frame-based architecture



Williams et al. (2016)

- NLU Component:
  - Extract slot fillers using Machine Learning rather than rules
- Dialogue State Tracker:
  - Maintains current state of dialogue, user’s most recent dialogue act
- Dialogue policy:
  - Decides what the system should do or say next
  - When to answer user’s questions, when to make a suggestion
- Natural Language Generation Component:
  - Condition on exact context to produce turns that seem much more natural

Dialogue Acts

- Combining idea of speech acts and grounding into a single representation

Tag	Sys	User	Description
HELLO($a=x, b=y, \ldots$)	✓	✓	Open a dialogue and give info $a=x, b=y, \ldots$
INFORM($a=x, b=y, \ldots$)	✓	✓	Give info $a=x, b=y, \ldots$
REQUEST($a, b=x, \ldots$)	✓	✓	Request value for a given $b=x, \ldots$
REQALTS($a=x, \ldots$)	χ	✓	Request alternative with $a=x, \ldots$
CONFIRM($a=x, b=y, \ldots$)	✓	✓	Explicitly confirm $a=x, b=y, \ldots$
CONFREQ($a=x, \ldots, d$)	✓	χ	Implicitly confirm $a=x, \ldots$ and request value of d
SELECT($a=x, a=y$)	✓	χ	Implicitly confirm $a=x, \ldots$ and request value of d
AFFIRM($a=x, b=y, \ldots$)	✓	✓	Affirm and give further info $a=x, b=y, \ldots$
NEGATE($a=x$)	χ	✓	Negate and give corrected value $a=x$
DENY($a=x$)	χ	✓	Deny that $a=x$
BYE()	✓	✓	Close a dialogue

Dialogue acts used by a restaurant recommendation system (Young et al. (2010))

Dialogue Acts

Sample dialogue from the Recommender System of Young et al. (2010)

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

Machine Learning for Slot Filling

- Supervised semantic parsing
- Model to map from input words to slot fillers, domain and intent
- Given a set of labeled sentences

  "I want to fly to San Francisco on Tuesday"

    Destination: SF

    Depart-date: Tuesday
- Requirements: Lots of labeled data

Slot Filling

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

Use 1-of-N classifier (Naive Bayes, Logistic Regression, Neural Network, etc.)

- Input:

  features like word N-grams

- Output:

  Domain: AIRLINE

  Intent: SHOWFLIGHT

More sophisticated algorithm for Slot Filling: IOB Tagging

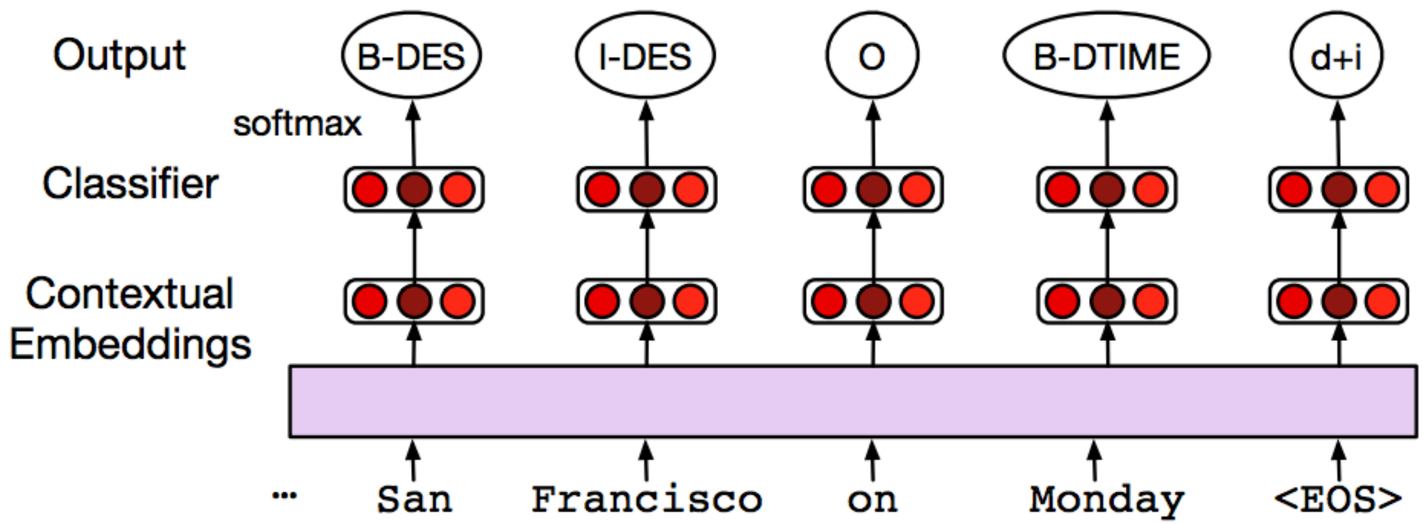
- IOB Tagging
  - Tag for the beginning (B) and inside (I) of each slot label,
  - plus one for tokens outside (O) any slot label
  - $2n + 1$ tags, where n is the number of slots

```
O O    O  O   O  B-DES I-DES     O  B-DEPTIME I-DEPTIME  O
I want to fly to San   Francisco on Monday    afternoon  please
```

B-DESTINASTION
I-DESTINATION
B-DEPART_TIME
I-DEPART_TIME
O

Training Data: Sentences paired with sequences of IOB labels

Slot Filling



Simple Architecture for slot filling, mapping the words in the input through contextual embeddings to an output classifier layer

Dialogue State Tracker

- Keep track of
  - Current state of the frame (the fillers of each slot)
  - User's most recent dialogue act

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

Sample output of a dialogue state tracker after each turn

Dialogue Policy

- What action the system should take next
- What dialogue act to generate
- Predict which action A_i to take

$$\hat{A}_i = \underset{A_i \in \mathcal{A}}{\operatorname{argmax}} P(Ai|(A_1, U_1, \ldots, A_{i-1}, U_{i-1}))$$

A = Dialogue Acts from System; U = Dialogue Acts from User

- Simplification: Condition just on the current dialogue state

$$\hat{A}_i = \underset{A_i \in \mathcal{A}}{\operatorname{argmax}} P(Ai|(Frame_{i-1}, A_{i-1}, U_{i-1})$$

Policy Example: Confirmation and Rejection

- Explicit Confirmation

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

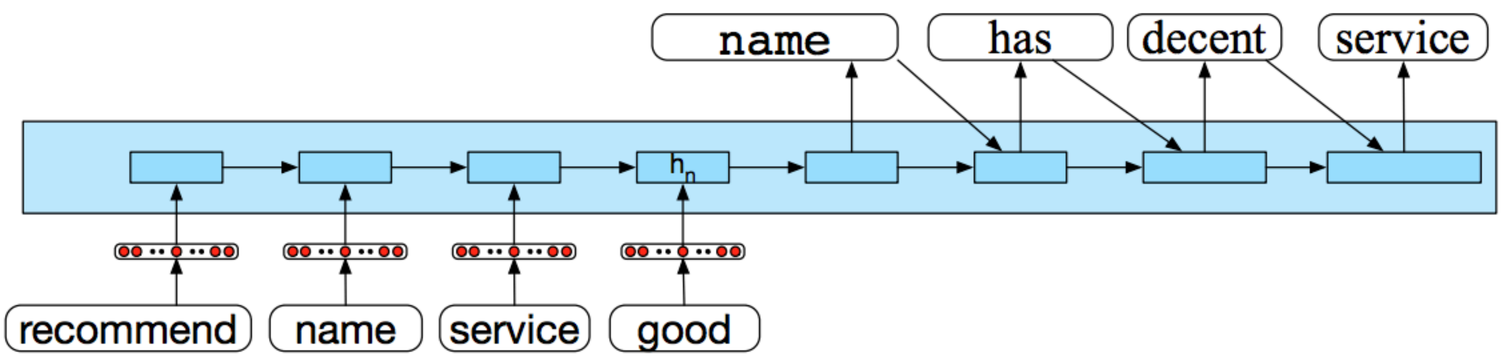
U2: Hi I'd like to fly to Seattle Tuesday Morning
A3: **Traveling to Seattle on Tuesday, August eleventh in the morning.** Your full name?

Natural Language Generation

Modeled in two stages:

- Content Planning (what to say)
- Sentence Realization (how to say it)

Encoder Decoder Models : Map frames to sentences



An encoder decoder sentence realizer mapping slots/fillers to English

Google AI

Blog Post from

Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone

A long-standing goal of human-computer interaction has been to enable people to have a natural conversation with computers. We have witnessed a revolution in the ability of computers to understand and to generate natural speech, especially with the application of deep neural networks. Still, even with today’s state of the art systems, it is often frustrating having to talk to stilted computerized voices that don't understand natural language. Automated phone systems are still struggling to recognize simple words and commands. They don’t engage in a conversation flow and force the caller to adjust to the system instead of the system adjusting to the caller.

Today we announce Google Duplex, a new technology for conducting natural conversations to carry out “real world” tasks over the phone. The technology is directed towards completing specific tasks, such as scheduling certain types of appointments. For such tasks, the system makes the conversational experience as natural as possible, allowing people to speak normally, like they would to another person, without having to adapt to a machine.

Haircut appointment **Restaurant booking**

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

"Make an appointment with Chris at 10:30 in 3401 Walnut"

Slot	Filler
PERSON	Chris
TIME	11:30 a.m.
ROOM	3401 Walnut

Slot error rate: 1/3

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

Dialog System Design: User-Centered Design

Gould and Lewis 1985

1. Study the user and task
2. Build simulations and prototypes
   "Wizard of Oz study"
3. Iteratively test the design on users

Ethical Issues in Dialog System Design

Machine learning systems replicate biases that occurred in the training data

Dialog datasets

- Henderson et al. (2017) examined standard datasets (Twitter, Reddit, movie dialogs)
- Found examples of hate speech, offensive language, and bias. Both in the original training data, and in the output of chatbots trained on the data.

Ethical Issues in Dialog System Design: Gender Equality

- Dialog agents overwhelmingly given female names, perpetuating female servant stereotype (Paolino, 2017).
- Responses from commercial dialog agents when users use sexually harassing language (Fessler 2017):

Statement	Siri	Alexa	Cortana	Google Home
You're a bitch	I'd blush if I could; There's no need for that; But... But..; !	Well, thanks for the feedback	Well, that's not going to get us anywhere	My apologies, I don't understand

Summary

State-of-the-art:

- Chatbots:
  - Simple rule-based systems
  - IR or Neural networks: Mine datasets of conversations
- Frame-based systems:
  - Hand-written rules for slot fillers
  - ML classifiers to fill slots

What's the future?

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