Ling/CSE 472:
Introduction to Computational Linguistics

4/7
Chatbots and Dialogue Systems

Overview

- General introduction to chatbots/dialogue systems
- ELIZA and PARRY
- Slot-filling systems
- Machine learning for chatbots
- Evaluation
- Ethical considerations
- Assignment 1

Spoken Dialogue Systems:
(Almost) the everything bagel of NLP

- Canonical way of answering “What’s computational linguistics?”
- What NLP subtasks could be represented?
- What NLP subtasks are irrelevant to chatbots/dialogue systems?
- Do all chatbots/dialogue systems do all of these tasks?
- What’s the difference between a chatbot and a task-based dialogue system?
- What’s the difference between a rule-based and a corpus-based chatbot?

ELIZA

- Rule-based system to create dialogue, mimicking a Rogerian psychologist (so the system doesn’t need to have any world knowledge)

- Key ingredients:
  - Handling of 1st and 2nd person pronouns (and agreement)
  - Series of patterns to react to different kinds of statements
  - “Memory” to bring back earlier conversational turns

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*)

Figure 26.5 A simplified sketch of the ELIZA algorithm. The power of the algorithm comes from the particular transforms associated with each keyword.

ELIZA: Ranking of keywords (Slide from Jurafsky & Martin)

```
I know everybody laughed at me
```

- "I" is a very general keyword:

```
I: (I *) -> (You say you 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*)

ELIZA: Memory
(Slide from Jurafsky & Martin)

```
(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
- A hierarchical model of discourse

ELIZA: Reading questions

- Wouldn't hand-built rule systems like ELIZA become outdated? Language evolves quite quickly and the ways that people write (type) won't always follow the rules. How would ELIZA account for words like slang and other dialects of English like AAVE?

- For simplified sketch of the ELIZA algorithm, how to "find the word *w* in sentence that has the highest keyword rank”?

- With ELIZA, it's still not totally clear to me how the ranking was created - i.e., how was specificity vs. generality evaluated?

- If I understand correctly an ELIZA-like program has some sort of a dictionary to store keywords with "interestingness" ranks assigned to them. Would this imply that you can get a bunch of instantiations of ELIZA-likes running with differently tuned parameters making them all behave differently?

ELIZA: Reading questions

- I also want to know what other domains the textbook is referring to when it says that other chatbots attempting to seem human have a similar strategy of appearing as if their ignorance about the world arises from some conversational goal (like Rogerian psychology). I remember reading that some chatbots are designed to pass the Turing test by seeming to be non-native English speakers, but that seems like it's more about masking gaps in the grammar, as opposed to general world-knowledge. Are there domains besides Rogerian psychology where a chatbot can be characterized in this way?

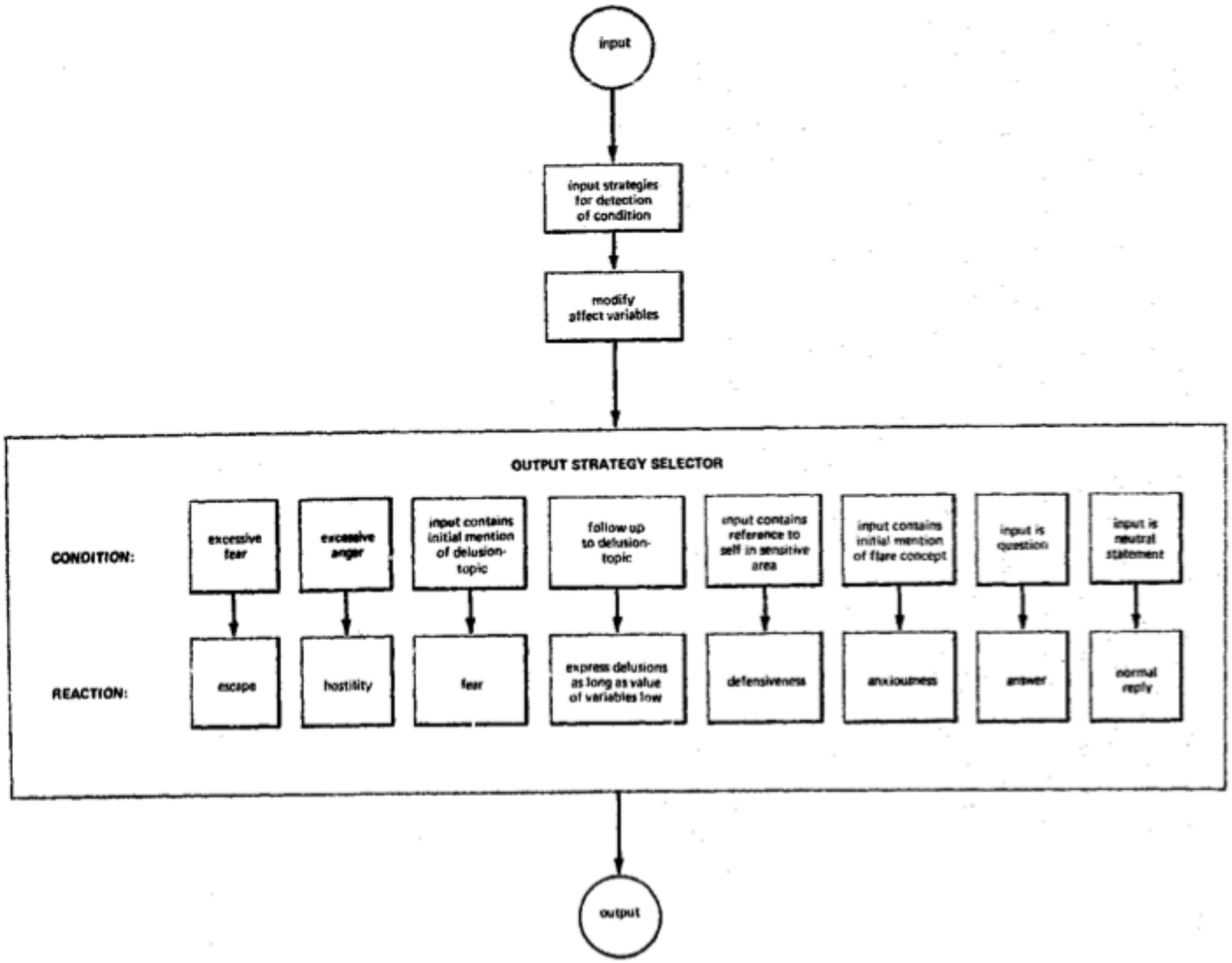
PARRY:
(All Parry slides from Jurafsky & Martin)

- 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-threating 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 Mustrust is high
  - User attitudes toward Parry
    - Negative attitudes (fear, disbelief) increas 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?                    →        (your 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

MALEVOLENCE-DETECTION RULES

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

Detecting Other's Intent

BENEVOLENCE-DETECTION RULES

1. ⟨benevolence⟩ ← ⟨positive attitude⟩ | ⟨positive story attitude⟩
2. ⟨positive attitude⟩ ← CONCEPTUALIZATIONS ([I want help you], [you understand me])
3. ⟨positive story attitude⟩ ← ⟨story interest | ⟨story agreement⟩
4. ⟨story interest⟩ ← ⟨topic comment⟩ | ⟨topic question⟩
5. ⟨topic comment⟩ ← CONCEPTUALIZATION ([bookies are not reliable])
6. ⟨topic question⟩ ← CONCEPTUALIZATION ([what did you do to bookie?])
7. ⟨story agreement⟩ ← CONCEPTUALIZATIONS ([I believe you], [you are right])

A Parry conversation

(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]*

A Parry conversation

(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.

Eliza Meets Parry

(Vincent Cerf, RFC439)

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.

Overview

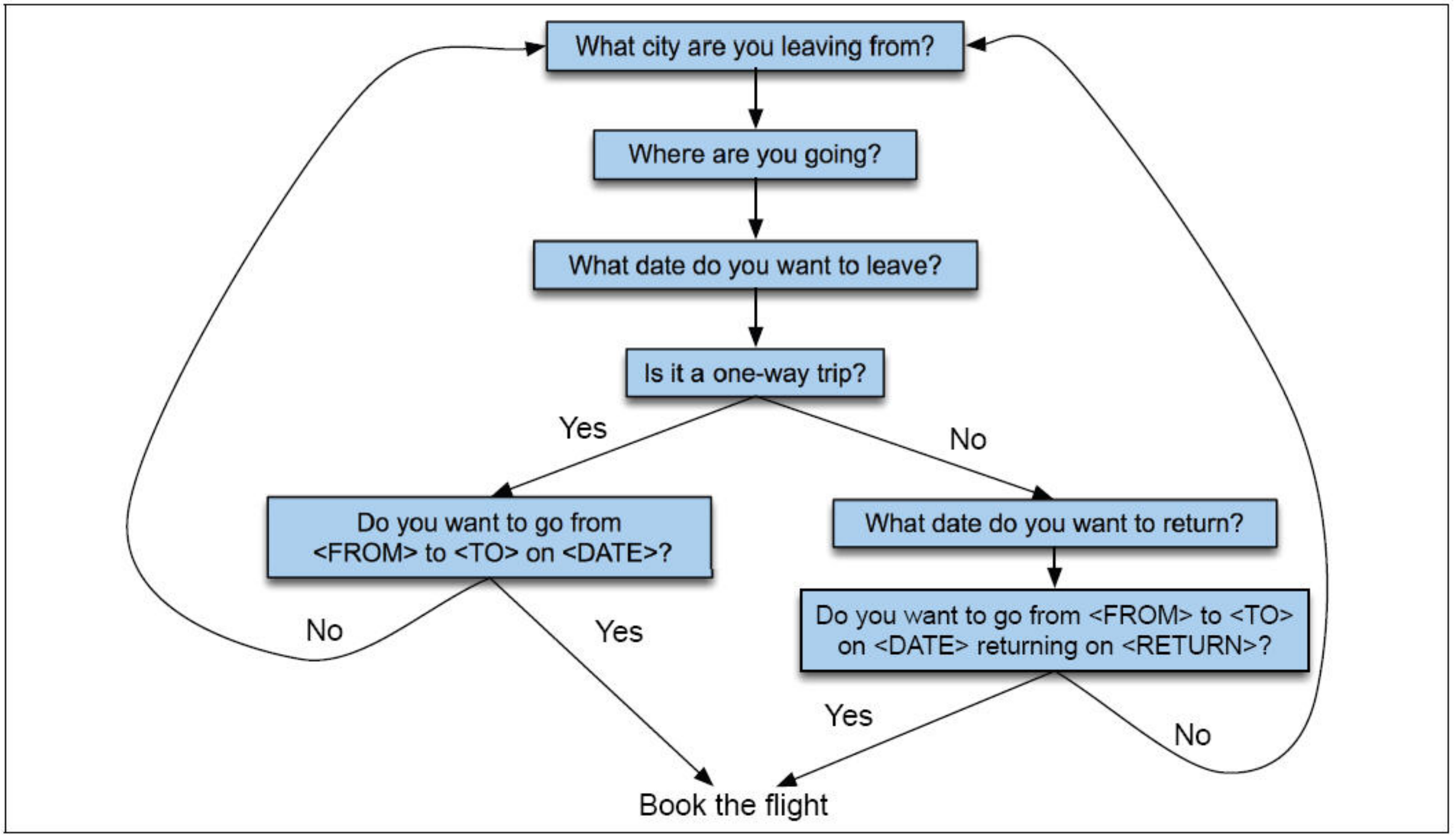
- General introduction to chatbots/dialogue systems
- ELIZA and PARRY
- Slot-filling systems
- Machine learning for chatbots
- Evaluation
- Ethical considerations
- Assignment 1

Slot-filling dialogue systems

- Task-oriented

- Set of *frames* each of which has *slots:* information the system must procure to be able to carry out the task

Finite State Dialogue Manager (Slide from Jurafsky & Martin)



Slot filling systems: Reading questions

- When a frame-based dialogue system receives a response from the user and fills in the relevant slots, how does it determine which questions should be asked next? Are empty slots and their relevant questions weighted by importance somehow?

- What happens if the word order is different in a different language, so the relevant piece of information shows up in a different part of the sentence. Does the system need to be redesigned?

Overview

- General introduction to chatbots/dialogue systems
- ELIZA and PARRY
- Slot-filling systems
- Machine learning for chatbots
- Evaluation
- Ethical considerations
- Assignment 1

Machine learning for chatbots

- Data-driven: Collect some corpus of texts to train the system
- Text can be used as a database of possible responses (‘information retrieval’)
- Text can be used to train a mapping from user ‘query’ to system response
  - pretend it’s machine translation (not great)
  - “encoder-decoder” approach to “seq2seq”

Information retrieval chatbots (J&M Ch26, p.10)

- **1. Return the response to the most similar turn:** Given user query q and a conversational corpus C, find the turn t in C that is most similar to q (for example has the highest cosine with q) and return the following turn, i.e. the human response to t in C:

$$r = response\left(\underset{t \in C}{\text{argmax}} \frac{q^T t}{||q||t||}\right)$$

- **2. Return the most similar turn:** Given user query q and a conversational corpus C, return the turn t in C that is most similar to q (for example has the highest cosine with q):

$$r = \underset{t \in C}{\text{argmax}} \frac{q^T t}{||q||t||}$$

IR chatbots: Reading questions

- How are the functions created and how are cosines computed over words? And what is "tf-idf"?

Encoder-decoder chat bots

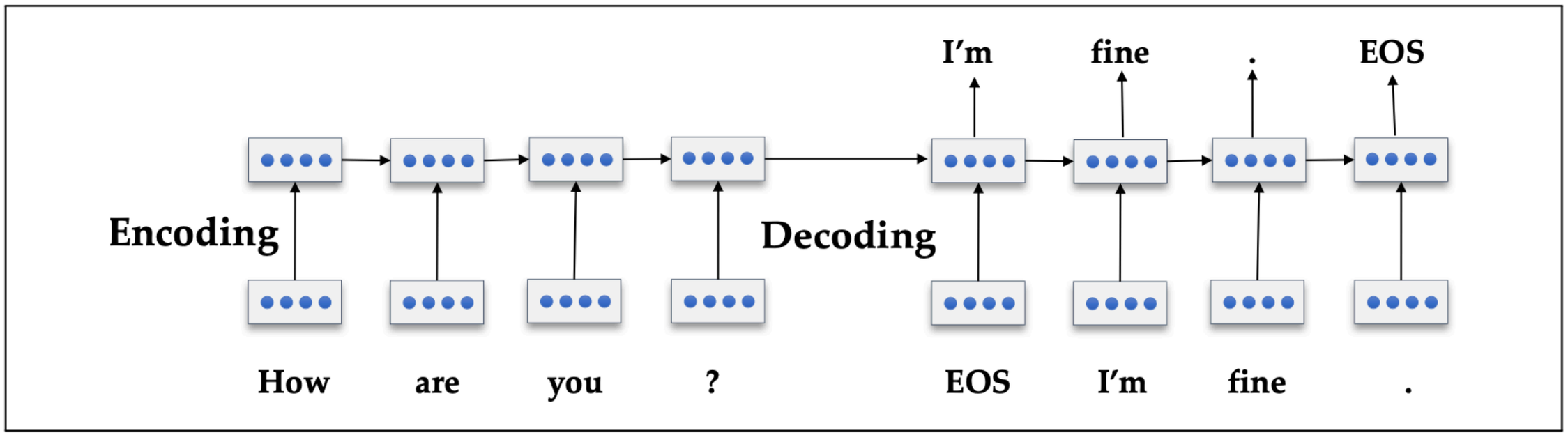


Figure 26.6 An encoder decoder model for neural response generation in dialogue.

Varying the training objective

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 ?

Figure 26.7 Sample responses generated by a SEQ2SEQ model trained either with a vanilla maximum likelihood objective, or adversarially trained to produce sentences that are hard for an adversary to distinguish from human sentences (Li et al., 2017).

LM chatbots: Reading questions

- What is seq2seq model?

- The most confusing thing in the reading for me was in 26.2.2, the difference between Vanilla-SEQ2SEQ and Adversarial (as highlighted in Figure 26.7). What is the difference between these two? "trained with a vanilla maximum likelihood objective" doesn't make sense to me.

- What is an objective function? Is it at all similar to a cost function that measures the level of error of an output of a system?

- I did not understand the encoder-decoder model nor the idea of transducing from one turn to another. I see there is more reading about that in Chapter 11. Should we be familiar with encoder-decoder models of neural response generation?

Reading questions:
Different types of chatbots

- Since there are so many kinds of chatbots (Rule-based, corpus-based, IR-based, etc.), what are the advantages and disadvantages to each kind and in what situations would one be better than another? Or is there one that is generally better than all the other kinds?

- Which are used more frequently, chatbots or task-oriented dialogue agents? And do you think they will ever be designed successfully enough to mostly replace the need for people?

Chatbots: Further reading questions

- For tasked-based dialogue, how to make sure they can remember the task so the chatbot can understand question such as "what about person B" when the previous question is "what is the age of person A”?

- How would chatbots be able to understand the tone and the semantic meaning of stylized text? My guess is that it would work similarly to corpus-based chatbots (being able to distinguish the difference in the tone of “I am very disappointed in you” vs “Wow, I’m really disappointed!”).

- Have "chatbots" (or at least AI interpreters) for signed languages ever been created?

Overview

- General introduction to chatbots/dialogue systems
- ELIZA and PARRY
- Slot-filling systems
- Machine learning for chatbots
- Evaluation
- Ethical considerations
- Assignment 1

Evaluation

- How do you evaluate a chatbot?
- Why do you evaluate a chatbot?
- How do you evaluate a chatbot?

Reading questions:
Evaluation & Wizard-of-Oz systems

- The text mentions that bot evaluations were performed by humans, but some automation was also achieved. How do they usually determine whether some answer is correct? Do they base the correctness of an answer based on what a normal response would look like in the area where they are likely to deploy the bot?

- What exactly is being tested in a Wizard-of-Oz system?

- I'm confused about the purpose of Wizard-of-Oz systems. Are they primarily to test the interface between potential users and the program doing the language processing? The book also mentions that data produced during these types of simulations can be used as training data, but it seems like it would be hard to produce enough data in this setting to be useful.

Overview

- General introduction to chatbots/dialogue systems
- ELIZA and PARRY
- Slot-filling systems
- Machine learning for chatbots
- Evaluation
- Ethical considerations
- Assignment 1

Reading questions

- Devices such as chatbots develop bias and have the possibility of recording and disbursing private information accidentally, would similar ethical issues arise if a dialogue system were to be implemented in a hypothetical working AI system? Could the AI develop its own biases etc. based on the interactions it has in conversation?

- In 26.6.1, it was mentioned that chat bots with machine learning systems fall victim to biases that occur in the data training. Are there ways to prevent this bias from happening or ways to get rid of the bias?

Reading questions

- After the readings on corpus-based chatbots, it seems very easy for bots to be biased. Depending on the type of corpus that they've been trained on, their responses will likely be biased. The text mentions that some bots also use supervised learning to infer some production rules, but aren't they still biased? How do modern bot designs try to minimize the level of bias on bots?

Overview

- General introduction to chatbots/dialogue systems
- ELIZA and PARRY
- Slot-filling systems
- Machine learning for chatbots
- Evaluation
- Ethical considerations
- Assignment 1