自然语言处理

对话系统

孙承杰 杨沐昀
sunchengjie@hit.edu.cn
哈尔滨工业大学计算学部
语言技术研究中心

主要内容

- 对话系统的应用
- 对话系统的分类
- 聊天型对话系统
  - 基于规则的的方法
  - 基于对话库的方法
    - 基于检索的聊天型对话系统
    - 基于生成的聊天型对话系统
- 任务型对话系统
  - 基于框架的任务型对话系统
- 对话系统的评价
- 对话系统的设计

对话系统的应用

- 手机上的个人助手（Personal Assistants）
  - SIRI, Alexa, Cortana, Google Assistant
- 临床应用于心理健康
- 闲聊、娱乐
- 车载对话系统
- 与机器人进行交流
- 完成特定任务，如订票
- 教育

对话系统的分类

- 聊天型 (Chatbots)
- 任务型(Goal-based or Task-oriented)
  - *interfaces to cars, robots…*
  - *booking flights or restaurants*

主要内容

- 对话系统的应用
- 对话系统的分类
- 聊天型对话系统
  - 基于规则的的方法
  - 基于对话库的方法
    - 基于检索的聊天型对话系统
    - 基于生成的聊天型对话系统
- 任务型对话系统
  - 基于框架的任务型对话系统
- 对话系统的评价
- 对话系统的设计

聊天型对话系统

- ELIZA (1966)
- PARRY (1971)
  - The first system to pass the Turing test!
- ALICE (1995)
- CLEVER (2006)
- Microsoft XiaoIce小冰 (2014)
- 小度 (百度，2017)
- 小爱同学 (小米，2017)
- 天猫精灵 (阿里，2017)
- Meena (Google, 2020)
- Blender (Facebook, 2020)

聊天型对话系统的架构

- 基于规则的 Rule-based
  - Pattern-action rules (Eliza)
  - + a mental model (Parry)

- 基于语料库的 Corpus-based (from large chat corpus)
  - 基于检索的 Information Retrieval (XiaoIce)
  - 基于生成的 Neural network encoder-decoder (Blender)

聊天型对话系统的功能

- 娱乐
  - 目前大多数聊天型机器人是娱乐性的
- 医疗
  - 临床治疗的测试平台（心理治疗、情感陪伴）

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 0F YOU?
My father
YOUR FATHER

Eliza: Weizenbaum (1966)

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

Eliza如何把天聊进行下去的?

- 谈话方式：假装对真实世界一无所知
  - 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”

Eliza聊天方式

Patient: "I went for a long boat ride”
Psychiatrist (Eliza模仿的对象): "Tell me about boats”

- You don’t assume she didn’t know what a boat is
- You assume she had some conversational goal
- Chatbots trying to pass the Turing test usually choose some such domain

Eliza pattern/transform rules

`(0 YOU 0 ME)` [*pattern*]

→

`(WHAT MAKES YOU THINK I 3 YOU)` [*transform*]

0 means Kleene *

The 3 is the constituent # in pattern

```
You hate me
WHAT MAKES YOU THINK I HATE YOU
```

Eliza的规则如何触发

keyword　　pattern　　Ranked list of transforms

I know everybody laughed at me

(I *) -> (You say you 2)

YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU

ELIZA的回复生成算法

```
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

`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*)

NONE（ELIZA的缺省回复）

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
- A 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

ELIZA的影响

- 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

PARRY: A computational model of schizophrenia （精神分裂症)

- Another chatbot with a clinical psychology focus
  - Colby, K. M., Weber, S., and Hilf, F. D. (1971). Artificial paranoia. *Artificial Intelligence 2*(1), 1–25.
- Used to study schizophrenia
- Same pattern-response structure as Eliza
- But a much richer:
  - control structure
  - language understanding capabilities
  - model of mental state.
    - variables modeling levels of Anger, Fear, Mistrust

Parry的人设

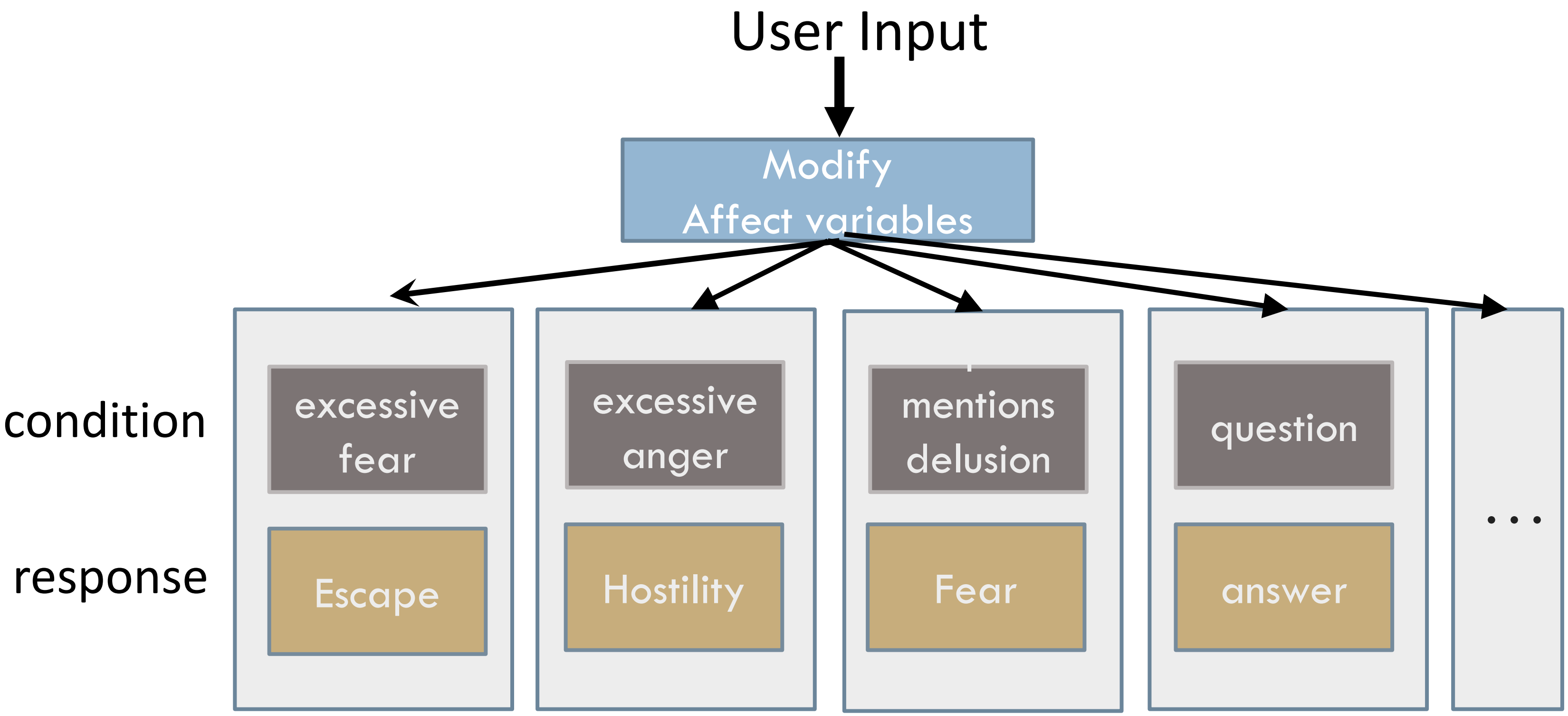
- 28-year-old single man, post office clerk
- no siblings and lives alone
- sensitive about his physical appearance, his family, his religion, his education and the topic of sex.
- hobbies are movies and gambling on horseracing,
- recently attacked a bookie, claiming the bookie did not pay off in a bet.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

Parry中的情感变量 Affect variables

- **Fear** (0-20) **Anger** (0-20) **Mistrust** (0-15)

- Start with all variables low
- After each user turn
  - Each user statement can change Fear and Anger
    - E.g., Insults increases Anger, Flattery decreases Anger
    - Mentions of his delusions increase Fear
  - Else if nothing malevolent in input
    - Anger, Fear, Mistrust all drop

Parry's responses depend on mental state



User Input

Modify
Affect variables

condition	excessive fear	excessive anger	mentions delusion	question	. . .
response	Escape	Hostility	Fear	answer	

PARRY passes the Turing test in 1972

- The first system to pass a version of the Turing test
- Psychiatrists couldn’t distinguish interviews with PARRY from (text transcripts of) interviews with people diagnosed with paranoid schizophrenia （偏执型分裂症)
  - Colby, K. M., Hilf, F. D., Weber, S., and Kraemer, H. C. (1972). Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. *Artificial Intelligence 3*, 199–221.

Parry系统对话示例

(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]*

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.

主要内容

- 对话系统的应用
- 对话系统的分类
- 聊天型对话系统
  - 基于规则的的方法
  - 基于对话库的方法
    - 基于检索的聊天型对话系统
    - 基于生成的聊天型对话系统
- 任务型对话系统
  - 基于框架的任务型对话系统
- 对话系统的评价
- 对话系统的设计

基于对话库的聊天型对话系统的两类方法

- Response by retrieval
  - Use information retrieval to grab a response (that is appropriate to the context) from some corpus
- Response by generation
  - Use a language model or encoder-decoder to generate the response given the dialogue context

对话语料库的来源?

- Transcripts of telephone conversations between volunteers
  - Switchboard corpus of American English telephone conversations
- Movie dialogue
  - Various corpora of movie subtitles
- Hire human crowdworkers to have conversations
  - Topical-Chat 11K crowdsourced conversations on 8 topics
  - EmpatheticDialogues 25K crowdsourced conversations grounded in a situation where a speaker was feeling a specific emotion
- Pseudo-conversations from public posts on social media
  - Drawn from Twitter, Reddit, Weibo (微博), etc.
  - Tend to be noisy; often used just as pre-training.
- Crucial to remove personally identifiable information

基于检索的聊天型对话系统

Idea: Mine conversations of human chats or human-machine chats

  - Microblogs: Twitter or Weibo (微博)
  - Movie dialogs

- Cleverbot http://www.cleverbot.com
- XiaoIce（小冰）https://www.xiaoice.com/
- Microsoft Tay

A Cleverbot conversation

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny
User:	What's your favorite show?
Cleverbot:	I don't have one
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User:	Because you said so.
Cleverbot:	I have to do my homework.

基于检索的聊天型对话系统的两种架构

- Return the response to the most similar turn
  - Take user's turn (*q*) and find a (tf-idf) similar turn *t* in the corpus C

    q = "do you like Doctor Who"

    t' = "do you like Doctor Strangelove"
  - Grab whatever the response was to *t*.

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

Yes, so funny

- Return the most similar turn

$$r = \operatorname*{argmax}_{t \in C} \frac{q^T t}{||q||t||}$$

Do you like Doctor Strangelove

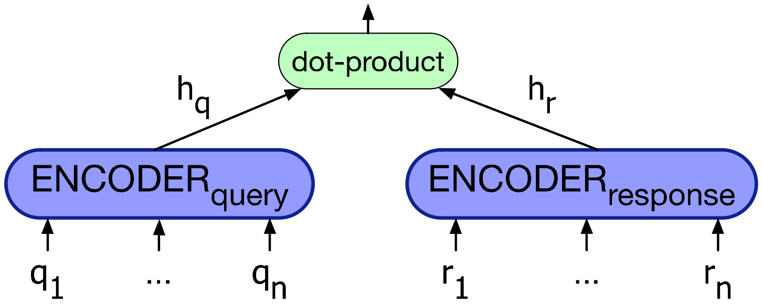
基于生成的聊天型对话系统

- Think of response generation as a task of *transducing* from the user’ s prior turn to the system’ s turn.
- Train on:
  - movie dialogue databases
  - Twitter conversations
- Train a deep neural network
  - map from user1 turn to user2 response

基于生成的聊天型对话系统

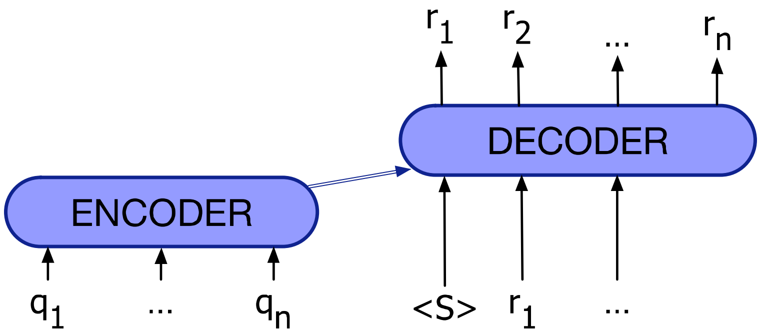
- Think of response production as an encoder-decoder task
- Generate each token r_t of the response by conditioning on the encoding of the entire query q and the response so far $r_1 ... r_{t-1}$

$$\hat{r}_t = \text{argmax}_{\text{w} \in \text{V}} P(w|q, r_1 ... r_{t-1})$$



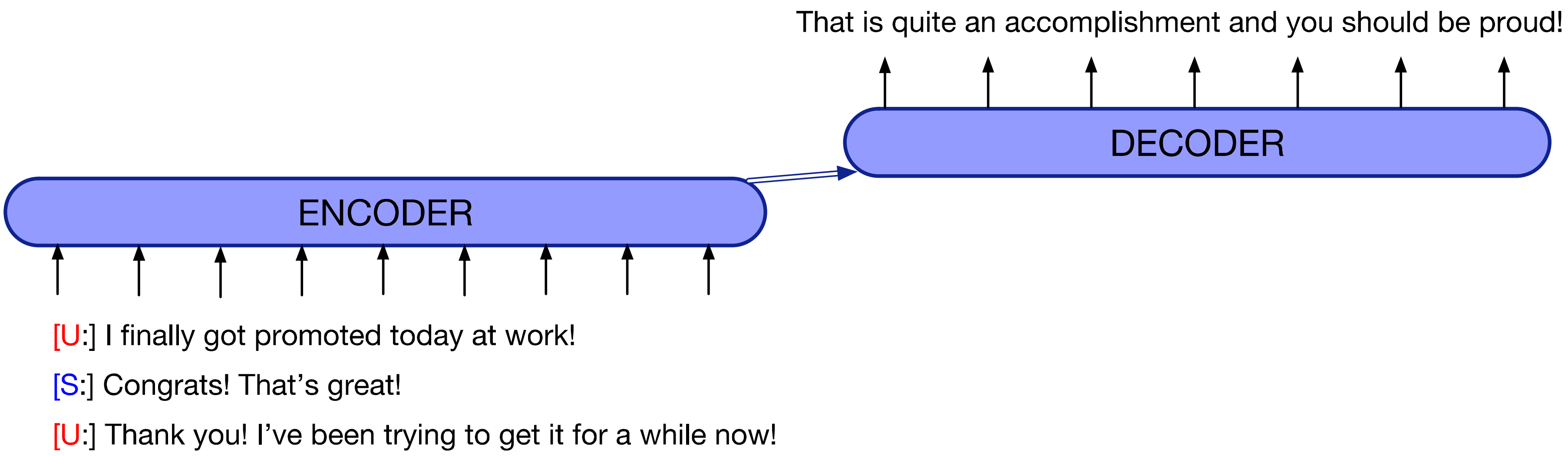
dot-product
h_q
h_r
$\text{ENCODER}_{\text{query}}$
$\text{ENCODER}_{\text{response}}$
q_1
...
q_n
r_1
...
r_n

(a) Response by Retrieval



(b) Response by Generation

基于生成的聊天型对话系统



[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!

基于生成的聊天型对话系统

- 存在的问题
  - 容易生成大量安全回复
  - 缺少知识
  - 无法保证一致性
  - 生成质量受训练数据影响大

聊天型对话系统的优缺点

- Pro:
  - Fun
  - Applications to counseling
  - 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 case of Microsoft Tay
      - Garbage-in, Garbage-out
- The future: combining chatbots with frame-based agents

主要内容

- 对话系统的应用
- 对话系统的分类
- 聊天型对话系统
  - 基于规则的的方法
  - 基于对话库的方法
    - 基于检索的聊天型对话系统
    - 基于深度学习的聊天型对话系统
- 任务型对话系统
  - 基于框架的任务型对话系统
- 对话系统的评价
- 对话系统的设计

基于框架的任务型对话系统

- 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:
  - First proposed in the GUS system of 1977
  - A knowledge structure representing user intentions
  - One or more **frames** (each consisting of **slots** with **values**)

框架（Frame)

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

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?

两种基本架构（本课程只讲第一种）

- The **GUS** architecture
  - Sometimes just called "**frame-based**" architecture
  - Over 40 years old, but still used in most industrial task-based dialogue agents
    - Bobrow, Daniel G., Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson, and Terry Winograd. 1977. "**GUS**, a frame-driven dialog system." *Artificial Intelligence* 8, 2:155-173.
- The **dialogue-state** architecture
  - Extension of GUS
  - More common in research systems
  - Some aspects making their way into industrial systems

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

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 the 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.

GUS: Natural Language Understanding for filling dialog slots

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

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

```
Wake me tomorrow at six.
```

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

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

How to fill slots?
(2) Machine learning for slot-filling

- Machine learning 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 words to semantic frame-fillers

- Requirements: Lots of labeled data

Machine learning for slot-filling: Domain and Intent

`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

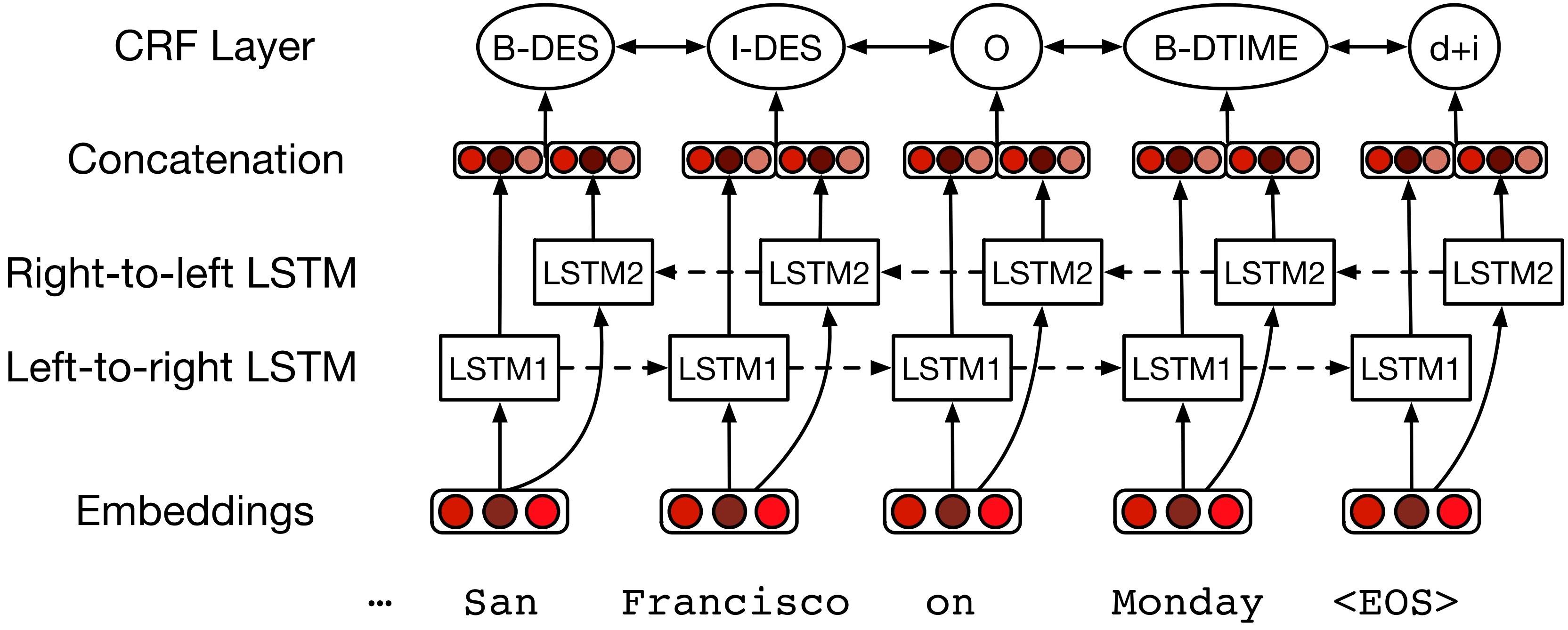
- IOB Tagging
  - tag for the beginning (B) and inside (I) of each slot label,
  - plus one for tokens outside (O) any slot label.
  - $2n + 1$ tags, where n is the number of slots.

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

```
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
```

More sophisticated algorithm for slot filling: IOB Tagging

- IOB Tagging is done by a sequence model
- Typical:



- Extracted strings can then be normalized (San Fran->SFO)

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?"

GUS架构总结: simple frame-based architecture

☐Like many rule-based approaches

- Positives:
  - High precision
  - Can provide coverage if the domain is narrow
- Negatives:
  - Can be expensive and slow to create rules
  - Can suffer from recall problems

主要内容

- 对话系统的应用
- 对话系统的分类
- 聊天型对话系统
  - 基于规则的的方法
  - 基于对话库的方法
    - 基于检索的聊天型对话系统
    - 基于深度学习的聊天型对话系统
- 任务型对话系统
  - 基于框架的任务型对话系统
- 对话系统的评价
- 对话系统的设计

对话系统的评价

- Chatbots:
  - mainly by human evaluation
- Task-based dialogue:
  - mainly by measuring task performance

Chatbots are evaluated by humans

- **Participant evaluation**: The human who talked to the chatbot assigns a score
- **Observer evaluation**: third party who reads a transcript of a human/chatbot conversation assigns a score.

Automatic evaluation is an open problem

- Automatic evaluation methods (like the BLEU scores used for Machine Translation) are generally not used for chatbots.
  - They correlate poorly with human judgements.
- One current research direction: **Adversarial Evaluation**
  - Inspired by the Turing Test
  - train a ``Turing-like'' classifier to distinguish between human responses and machine responses.
  - The more successful a dialogue system is at fooling the evaluator, the better the system.

Task-based systems are evaluated by task success!

- End-to-end evaluation (Task Success)

- Slot Error Rate for a Sentence

$$\frac{\text{\# of inserted/deleted/substituted slots}}{\text{\# of total reference slots for sentence}}$$

Evaluation Metrics: Slot error rate

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

Slot	Filler
PERSON	Chris
TIME	11:30 a.m.
ROOM	Gates 104

Slot error rate: 1/3

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

主要内容

- 对话系统的应用
- 对话系统的分类
- 聊天型对话系统
  - 基于规则的的方法
  - 基于对话库的方法
    - 基于检索的聊天型对话系统
    - 基于生成的聊天型对话系统
- 任务型对话系统
  - 基于框架的任务型对话系统
- 对话系统的评价
- 对话系统的设计

Dialog System Design: User-centered Design

Gould and Lewis 1985

1. Study the user and task
2. Build simulations

   "Wizard of Oz study"
3. Iteratively test the design on users

Gould, John D., and Clayton Lewis. "Designing for usability: key principles and what designers think." *Communications of the ACM* 28, no. 3 (1985): 300-311.

Bender, Emily M., and Batya Friedman. "Data statements for natural language processing: Toward mitigating system bias and enabling better science." TACL 6 (2018): 587-604.

Ethical Issues in Dialog System Design

- Ethical issues have long been known to be crucial in artificial agents
- Ethical issues:
  - **Safety**: Systems abusing users, distracting drivers, or giving bad medical advice
  - **Representational harm**: Systems demeaning particular social groups
  - **Privacy**: Information Leakage

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

Peter Henderson, Koustuv Sinha, Nicolas Angelard-Gontier, Nan Rosemary Ke, Genevieve Fried, Ryan Lowe, and Joelle Pineau. 2018. Ethical Challenges in Data-Driven Dialogue Systems. In 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’18),

Abuse and Representation Harm: 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
  - used informal language, slang, emojis, and GIFs,
  - Designed to learn from users (IR-based)

Abuse and Representation Harm: The case of Microsoft Tay

- Immediately Tay turned offensive and abusive
  - Obscene and inflammatory tweets
  - Nazi propaganda, conspiracy theories
  - Began harassing women online
  - Reflecting racism and misogyny of Twitter users
- Microsoft took Tay down after 16 hours
- Lessons:
  - User response must be considered in the design phase

Gina Neff and Peter Nagy 2016. Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication* 10(2016), 4915–4931

Privacy

- Remember this was noticed in the days of Weizenbaum
- Agents may record sensitive data
  - (e.g. “Computer, turn on the lights [answers the phone –Hi, yes, my password is…”],
- Which may then be used to train a seq2seq conversational model.
- Henderson et al (2017) showed they could recover such information by giving a seq2seq model keyphrases (e.g., "password is")

总结

- 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

本章参考阅读

- Chatbots and Dialogue Systems
  - https://web.stanford.edu/~jurafsky/slp3/24.pdf
  - https://web.stanford.edu/~jurafsky/slp3/slides/24_Dialogue_May_6_2021.pptx