Computational Dialogue Modelling

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam

NLP1 2018 (Guest Lecture)

Conversation

Primary setting for language use

- *multi-agent*: requires coordination (joint action)
- *spontaneous and online*: disfluent, fragmentary

A transcript fragment from the Switchboard corpus:

```
B.52 utt1:  Yeah,
B.52 utt2:  it's - it's fun getting together with immediate family.
B.52 utt3:  A lot of my cousins are real close
B.52 utt4:  and we always get together during holidays and
             weddings and stuff like that,
A.53 utt1:  Uh, those are the ones that are in Texas?
B.54 utt1:  # Uh, no #
A.55 utt1:  # Or you # go to Indiana on that?
B.56 utt1:  the ones in Indiana,
B.56 utt2:  uh-huh.
A.57 utt1:  Uh-huh,
A.57 utt2:  where in Indiana?
B.58 utt1:  Lafayette.
A.59 utt1:  Lafayette, I don't know where,
A.59 utt2:  I used to live in Indianapolis.
B.60 utt1:  Yeah,
B.60 utt2:  it's a little north of Indianapolis, about an hour.
```

Dialogue Modelling

What?

- the conversation from outsider's point of view, to retrieve information (summarisation, etc)
- the capabilities required to take part in a conversation
  – model a dialogue agent → *focus today*

Why?

- scientific motivation: gain understanding on human dialogue abilities
- technological motivation: develop dialogue systems that are useful
- *both!*

How?

- *we'll see different approaches today*

First chatbot: Eliza (Weizenbaum, 1966)

A conversation with Eliza:

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

(Jurafsky & Martin, SLP-3)

What? Given the dialogue partner's utterance, provide a response
Why? Psychological therapist / Turing test

First chatbot: Eliza (Weizenbaum, 1966)

How? Pattern / transformation rules based on keywords

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?
```

(Jurafsky & Martin, SLP-3)

Data-driven chatbots

From rule-based to *data-driven chatbots*: exploit large corpora of conversations (movie transcripts, Twitter, etc.)

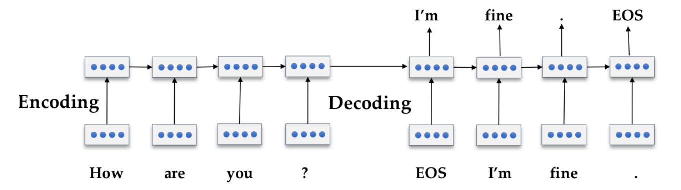
What? Given the dialogue partner's utterance, provide a response
Why? Social chit-chat
How?

- *Information Retrieval*
  retrieve a response from a corpus
  (e.g. return response to most similar turn in corpus)
- *Generation*
  generate a response exploiting corpus-based information
  $\rightarrow$ neural network encoder-decoder models

Encoder-Decoder models

Two main components, an encoder RNN and a decoder RNN:

- **Encoder:** converts the input into a fix-size feature vector that represents important information.
- **Decoder:** generates (word-by-word) the response with the highest probability in the training corpus given the input.



(Jurafsky & Martin, SLP-3)

Evaluation: string matching measures on human response in test data.

Encoder-Decoder models

Examples of generated responses:

U:	tell me ... how long have you had this falling sickness ?
S:	i'm not a doctor.
U:	How long does it take you guys to learn the drill?
S:	Like 2 weeks ago!!
U:	Where do you live?
S:	I live in England.
U:	In which city do you live?
S:	I live in Los Angeles.

[adapted from Jurafsky & Martin, SLP-3]

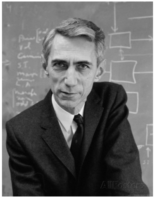
Problems:

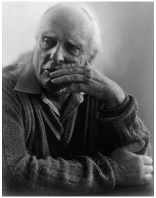
- no real understanding, lack of coherence / consistency
- reproducing general surface patters in the training data

⇝ *dialogue happens in context – there is more than what's on the surface of an utterance*

Two views of communication

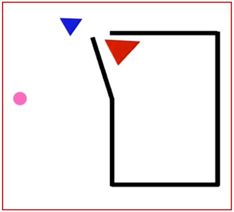
Shannon (1948) – Information theory: communication as information encoded by a sender and decoded by a recipient.





Grice (1957): human communication is characterised by the process of *intention recognition*

Goals and intentions



original video

Heider & Simmel (1944): An experimental study in apparent behaviour

- tendency to ascribe *goals* and *intentions*
- theory of mind: ability to model internal mental state of agents
- attribution of causation

Any sensing actions, including linguistic actions, trigger the attribution of mental attitudes and goals

- *Speech act theory:* conversations are made up of *linguistic actions*.

Speech Act Theory

Initiated by Austin (*'How to do things with words'*) and developed by Searle in the 60s-70s within philosophy of language.

Examples of *dialogue acts*: *inform, apologise, promise, command, request, answer, . . .*

- The director bought a new car this year.
- Sorry for being late.
- I'll surely come to your talk tomorrow afternoon.
- Put the car in the garage, please.
- Is she a vegetarian?

On the Gricean view, it is possible for the same surface form to correspond to different intentions:

The gun is loaded $\rightsquigarrow$ *threatening? warning? explaining?*

Also, the same intention can be realised by different utterances.

Task-Oriented Dialogue Systems

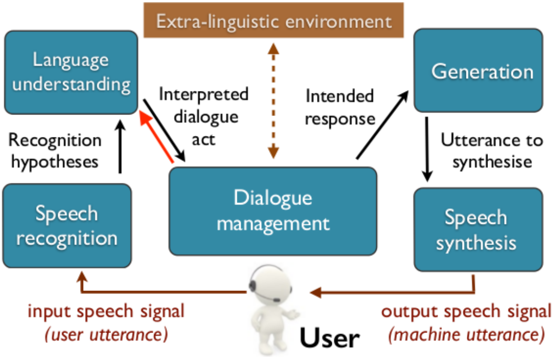
- Dialogue acts capture goals and intentions of the participants.
- They are a better clue for how to respond in dialogue than simply surface form.

Task-oriented dialogue systems:

- a task / end goal allows us to make intentions tractable
- more reliable evaluation
- more useful systems that help us accomplish goals

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

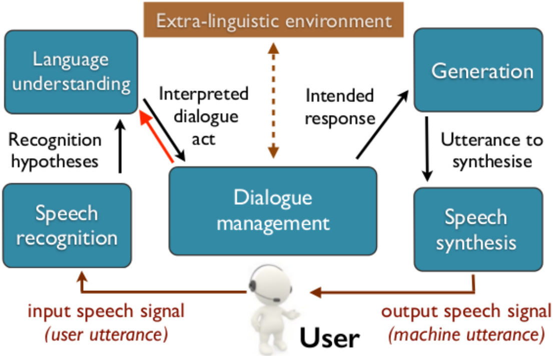
Modular Dialogue System Architecture



Language understanding: the NLP1 course!

- morphological processing, POS tagging
- Lexical semantics
- Syntactic parsing
- Compositional semantics

Modular Dialogue System Architecture

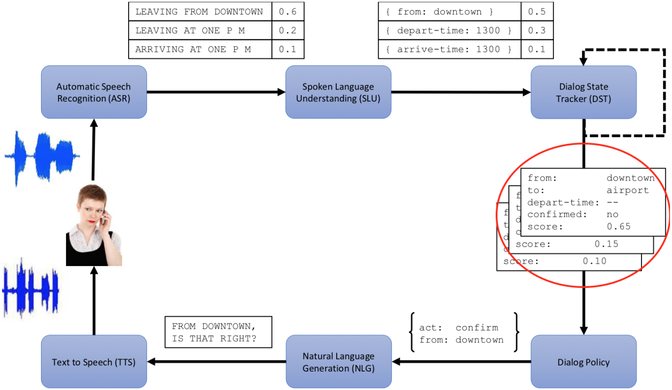


Dialogue Management: two main components

- *Dialogue state tracker:* linguistic context (what has been said) and how this is relevant for the task at hand
- *Dialogue policy:* next action selection (what to say next)

Modular Dialogue System Architecture

Consider a travel domain: The *dialogue state* can be modelled as a frame with task-related slots that need to be filled in.



Dialogue State Tracker

Dialogue acts are defined relative to a task/domain:

	Utterance	Dialog 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()

(Jurafsky & Martin, SLP-3)

Dialogue State Tracker

The state tracker needs to interpret the latest dialogue act and integrate it into the state:

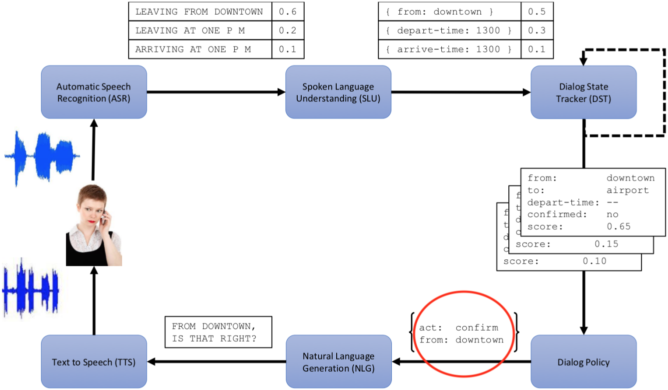
User:	I'm looking for a cheaper restaurant
	`inform(price=cheap)`
System:	Sure. What kind - and where?
User:	Thai food, somewhere downtown
	`inform(price=cheap, food=Thai, area=centre)`
System:	The House serves cheap Thai food
User:	Where is it?
	`inform(price=cheap, food=Thai, area=centre); request(address)`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

(Jurafsky & Martin, SLP-3)

- *Dialogue act interpretation* can be modelled as a supervised classification task (with feature-based or neural classifier)
- *Slot filling* can be modelled as supervised sequence tagging: assign a slot value to each word in the utterance.

Modular Dialogue System Architecture

The goal of the *dialogue policy* is to decide what action the system should take next: what dialogue act to generate.



Dialogue Policy

We can condition our decision on the current dialogue state

(abstraction over entire history: different dialogues could lead to the same state)

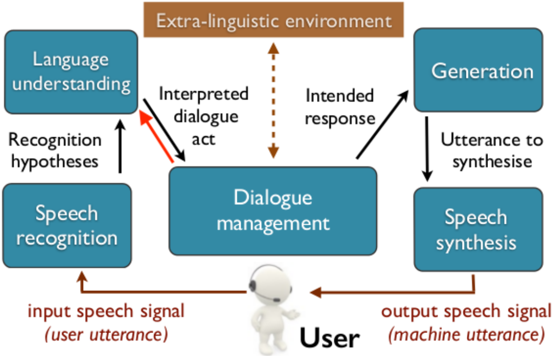
$$A_t = \underset{A_i \in A}{\text{argmax}}\, P(A_i|\text{Frame}_{t-1}, A_{t-1}, U_{t-1})$$

- Frame: current dialogue state (filled slots so far)
- A_{t-1}: latest action by the system
- U_{t-1}: latest dialogue act by the user
- A: set of available system actions

These probabilities can be estimated from large corpora of annotated conversations (often simulations are needed).

→ Reinforcement Learning has been used to select actions that are likely to lead to task success.

Modular Dialogue System Architecture



Extra-linguistic environment: different options, depending on the type of system

- Database for the domain at hand or/and world knowledge
- Perceptual environment, for example modelled by an image
  → more in this direction by Elia Bruni later

Summing Up

- Open-domain *chatbots* are fun, but they current systems miss out on key properties of conversation, are difficult to evaluate, and are only relatively useful.
- Classic *modular task-oriented systems* are potentially useful and capture key properties of conversation, but require large amounts of annotated data.

- Future: task-oriented systems that learn their own representations end-to-end, with no manual annotation.
- See further reading (tutorial at COLING 2018 and references therein) for the latest developments.