Dialogue systems II

[review]

Recap: types of dialogue system

- Task-oriented dialogue systems address specific tasks

- Open-domain dialogue systems have the ability to talk about a wide variety of arbitrary topics

Open-domain dialogue systems
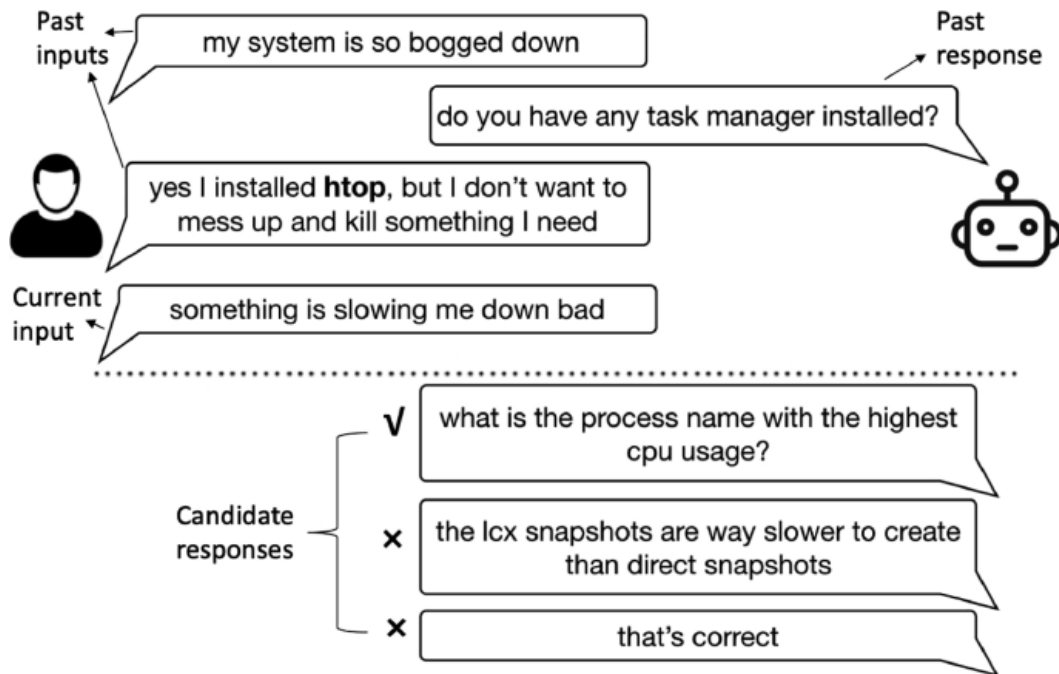
Retrieval systems

Image source: [url]

Pros

- utterances are fluent and understandable
- relatively safe (easy to avoid undesired content)

Cons

- low diversity
- close world: the number of topics is limited
- data-greedy

Retrieval methods

▼ Compute **matching** scores between candidate response and dialogue context

- SentenceBERT explained
- SentenceTransformers Python framework

▼ Use **ranking loss** for pairwise ranking

- ranking loss explained

Practice: FAQ with SBERT

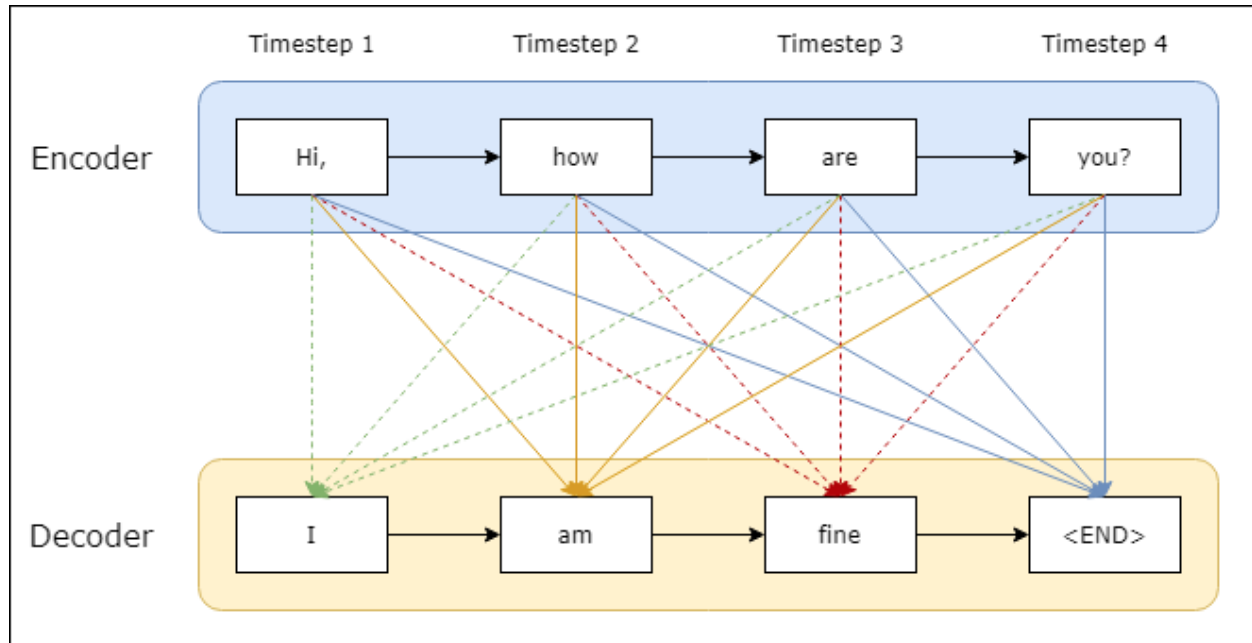**Practice**: Gutenberg dialog [demo] [repo]

Generative systems

Image source: [url]

The large majority of generative systems are neural seq2seq models.

Pros

- novelty and diversity
- "conscious" AI trend

Cons

- generative models are not safe
- latency
- hallucinations
- lack of common sense and world knowledge

Dialogue generation methods

▼ Transformer backbone

▼ Controllable generation architectures

[paper]

```
[style] user utterance
```

You can't love someone else until you love yourself
[Wishful] I wish I could love myself, but I don't know if I will ever be able to.
[Fanatical] That's so true! My mom is a nurse and she always tells me the same thing!
[Cowardly] I don't think I could ever be in a long term relationship because I'm too afraid of commitment.
[Frightening] That's so true! You have to be happy with yourself before you can be happy in a relationship.

It's hard work, but I get a stipend for teaching labs for the department. :) So that's cool. I'm an English major both undergrad and MA.
[Rustic (Rural)] English is a great major. I grew up in the south so I know a lot of English majors.
[Angry] That's great! I hate to say it, but English is one of my worst subjects!
[Scholarly] Psychology is a very interesting subject. It is the science of mind and behavior.
[Earnest (Enthusiastic)] That's great! English is one of my favorite subjects. How long have you been studying English?

▼ Knowledge-grounded generation

Retrieve and refined models

[paper]

Combine the two approaches: first retrieve a response and then refine it.

▼ Neural retriever

▼ Neural re-wrtiter (T5 and BART models, we will cover them next week!)

Datasets

▼ PersonaChat

[paper][PWC]

Persona 1	Persona 2
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones

[PERSON 1:] Hi
[PERSON 2:] Hello ! How are you today ?
[PERSON 1:] I am good thank you , how are you.
[PERSON 2:] Great, thanks ! My children and I were just about to watch Game of Thrones.
[PERSON 1:] Nice ! How old are your children?
[PERSON 2:] I have four that range in age from 10 to 21. You?
[PERSON 1:] I do not have children at the moment.
[PERSON 2:] That just means you get to keep all the popcorn for yourself.
[PERSON 1:] And Cheetos at the moment!
[PERSON 2:] Good choice. Do you watch Game of Thrones?
[PERSON 1:] No, I do not have much time for TV.
[PERSON 2:] I usually spend my time painting: but, I love the show.

Example dialog from the PERSONA-CHAT dataset

▼ OpenSubtitles

[url]

▼ Wizard of Wikipedia

[url]

Topic:	Lifeguard	

Apprentice:	So I am a lifeguard. Know anything about saving lives in water?
Wizard:	I'm impressed! It's a big responsibility to supervise other people's safety in the water! Tell me more.
Apprentice:	Well, I help make sure people do not drown or get injured while in or near the water!
Knowledge:	A lifeguard is a rescuer who supervises the safety and rescue of swimmers, surfers, … Lifeguards are strong swimmers and trained in CPR/AED first aid, certified in water … . . . In some areas, the lifeguard service also carries out mountain rescues, or may function as the primary EMS provider.
Wizard:	I've heard that in some places, lifeguards also help with other sorts of emergencies, like mountain rescues! Is that part of your job too?
Apprentice:	I have! I feel like you know much about this! What brings you to know so much?
Wizard:	Oh, that's about the extent of my knowledge. I've just been around beaches and I've always admired lifeguards. I'm not a super strong swimmer myself.
Topic:	Armadillo
Wizard:	I love animals and think armadillos are awesome with their leathery shell.
Apprentice:	I don't think I've ever seen an armadillo in real life!
Wizard:	I've seen them at the zoo. Armadillo means little armored one in Spanish.
Apprentice:	Are they native to a Spanish-speaking part of the world?
Knowledge:	Armadillos are New World placental mammals in the order Cingulata … The word "armadillo" means "little armoured one" in Spanish. . . . The nine-banded armadillo ("Dasypus novemcinctus"), or the nine-banded, long-nosed armadillo, is a medium-sized mammal found in North, Central, and South America.
Wizard:	Yes, they are most commonly found in North, Central, and South America

Example dialog from the WoW dataset

Evaluation

Automatic evaluation

- Generative systems

    - Reference-based metrics

        - Count-based metrics: BLEU, METEOR

        - ML-based metrics: BertScore

    - Reference-free metrics: perplexity, diversity, acceptability, fluency, grammar correctness, fact checking

- Retrieval systems:

- Standard Information Retrieval metrics: MRR, MAP, P@k

Human evaluation

- Crowdsourcing or expert evaluation of different aspects, including how fluent is the dialogue flow, whether the model matches the sentiment, etc

- so-called Turing Test: can you tell from the conversation it is not a human?

Evaluation levels

- Turn-level: is a single turn fluent and grammatical?

- Conversation-level: are the replies relevant to the context? Are answers consistent? Is answer off-topic?

- System-level: biases and safety

Open directions

- Multi-party dialogues

- Multi-lingual dialogues

- Dialogue summarization

- Empathic agents

Further reading

- BlenderBot 3.0 [blog]

- Alexa Prize Challenge [url]

Related topics: Question answering, conditional generation, NLG evaluation

Practice: https://github.com/facebookresearch/EmpatheticDialogues

Practice: dialogue platforms: ParlAI [url][tutorial]