Lecture 16: Dialogue

Alan Ritter

(many slides from Greg Durrett)
This Lecture

- Chatbot dialogue systems
- Task-oriented dialogue
- Other dialogue applications
Chatbots
Turing Test (1950)

- Imitation game: A and B are locked in rooms and answer C’s questions via typewriter. Both are trying to act like B
Turing Test (1950)

- Imitation game: A and B are locked in rooms and answer C’s questions via typewriter. Both are trying to act like B

  Original Interpretation:
Turing Test (1950)

- Imitation game: A and B are locked in rooms and answer C’s questions via typewriter. Both are trying to act like B

Original Interpretation:

```
A  B
-----
C  trained judge
```
Imitation game: A and B are locked in rooms and answer C’s questions via typewriter. Both are trying to act like B.
Turing Test (1950)

- Imitation game: A and B are locked in rooms and answer C’s questions via typewriter. Both are trying to act like B

Original Interpretation:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C trained judge</td>
<td>C trained judge</td>
</tr>
</tbody>
</table>

Standard Interpretation:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C trained judge</td>
<td>C trained judge</td>
</tr>
</tbody>
</table>
Turing Test (1950)

- Imitation game: A and B are locked in rooms and answer C’s questions via typewriter. Both are trying to act like B.

Original Interpretation:

- C (trained judge) observes conversations between A and B.

Standard Interpretation:

- The test is not “does this computer seem human-like to random people with a web browser?”
ELIZA

- Created 1964-1966 at MIT, heavily scripted
- DOCTOR script was most successful: repeats user’s input, asks inane questions

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 VERY 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 AM 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

Weizenbaum (1966)
ELIZA

- Identify keyword, identify context, apply transformation rule
Identify keyword, identify context, apply transformation rule

\[(.*)\) you \[(.*)\] me\]

Why do you think I $2 you?

You don’t argue with me.

**WHY DO YOU THINK I DON’T ARGUE WITH YOU**
ELIZA

- Identify keyword, identify context, apply transformation rule

\[
\begin{align*}
(\cdot\cdot) \text{ you (\cdot\cdot) me} & \quad \Rightarrow \quad \text{Why do you think I $2 you?} \\
\text{My (\cdot) (\cdot\cdot)} & \quad \Rightarrow \quad \text{What else comes to mind when you think of your $1?}
\end{align*}
\]

You don’t argue with me.

WHY DO YOU THINK I DON’T ARGUE WITH YOU
Identify keyword, identify context, apply transformation rule

\[(.*)\) you \((.*)\) me \n\rightarrow \nWhy do you think I $2 you?\]

\[My \((.*)\) \((.*)\) \n\rightarrow \nWhat else comes to mind when you think of your $1?\]

You don’t argue with me.
WHY DO YOU THINK I DON’T ARGUE WITH YOU

Very little need to generate new content, but can only have one type of conversation
Cleverbot

- Carpenter (1986), online system built in 2006

User: Hi there cleverbot.
Cleverbot: How are you?
User: Doing okay. what plans do you have today?
Cleverbot: I do revision.
User: What are you revising for?
Cleverbot: Maths history.
User: How old are you?
Cleverbot: That's none of your business.
User: What is maths history?
Cleverbot: I don’t know, farming maybe
Cleverbot

- Carpenter (1986), online system built in 2006
- “Nearest neighbors”: human says statement A, find a human response in human-human or human-computer chats to statement A, repeat that

User:
Hi there cleverbot.
Cleverbot: How are you?
User: Doing okay. what plans do you have today?
Cleverbot: I do revision.
User: What are you revising for?
Cleverbot: Maths history.
User: How old are you?
Cleverbot: That's none of your business.
User: What is maths history?
Cleverbot: I don’t know, farming maybe
Cleverbot

- Carpenter (1986), online system built in 2006
- “Nearest neighbors”: human says statement A, find a human response in human-human or human-computer chats to statement A, repeat that

- Can often give sensible answers, but the bot doesn’t really impose high-level discourse structure

User: Hi there cleverbot.
Cleverbot: How are you?
User: Doing okay. what plans do you have today?
Cleverbot: I do revision.
User: What are you revising for?
Cleverbot: Maths history.
User: How old are you?
Cleverbot: That's none of your business.
User: What is maths history?
Cleverbot: I don’t know, farming maybe
Data-Driven Approaches

- Can treat as a machine translation problem: “translate” from current utterance to next one

Ritter et al. (2011)
Data-Driven Approaches

- Can treat as a machine translation problem: “translate” from current utterance to next one

Ritter et al. (2011)
Data-Driven Approaches

- Can treat as a machine translation problem: “translate” from current utterance to next one

- Filter the data, use statistical measures to prune extracted phrases to get better performance

Ritter et al. (2011)
## Data-Driven Approaches

<table>
<thead>
<tr>
<th>Status</th>
<th>Mt-Chat</th>
<th>Mt-Baseline</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>riding to the arc on my bike. Hopefully I won’t melt into the pavement.</td>
<td>good luck on your bike ride arc</td>
<td>i can’t wait for it! i’m arc is a year.</td>
<td>make sure to ride in the middle of the street</td>
</tr>
<tr>
<td>I really hate watching movies in the living room.</td>
<td>me too, what movies are you doing in room?</td>
<td>me too! i love those? ) in my mouth.</td>
<td>Why? no comfy spots or just too open to public</td>
</tr>
<tr>
<td>Hello Did I miss David Getting his teen choice award? Or did they not show it:(</td>
<td>hello miss? * he was getting his will probably win.</td>
<td>it’s just like what you just did the exact date hello?</td>
<td>nothing yet...he presented though he is so darn cute</td>
</tr>
</tbody>
</table>

Ritter et al. (2011)
 Seq2seq models

- Just like conventional MT, can train seq2seq models for this task
Just like conventional MT, can train seq2seq models for this task

Why might this model perform poorly? What might it be bad at?
Seq2seq models

- Just like conventional MT, can train seq2seq models for this task
- Why might this model perform poorly? What might it be bad at?

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANDOM</td>
<td>0.33</td>
</tr>
<tr>
<td>MT</td>
<td>3.21</td>
</tr>
<tr>
<td>HUMAN</td>
<td>6.08</td>
</tr>
</tbody>
</table>
Lack of Diversity

Training to maximize likelihood gives a system that prefers common responses:

<table>
<thead>
<tr>
<th>Input: What are you doing?</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.86 I don’t know.</td>
</tr>
<tr>
<td>-1.03 I don’t know!</td>
</tr>
<tr>
<td>-1.06 Nothing.</td>
</tr>
<tr>
<td>-1.09 Get out of the way.</td>
</tr>
<tr>
<td>-1.09 I’m going home.</td>
</tr>
<tr>
<td>-1.09 Oh my god!</td>
</tr>
<tr>
<td>-1.10 I’m talking to you.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input: what is your name?</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.91 I don’t know.</td>
</tr>
<tr>
<td>-0.92 I don’t know!</td>
</tr>
<tr>
<td>-0.92 I don’t know, sir.</td>
</tr>
<tr>
<td>-0.97 Oh, my god!</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>-1.55 My name is Robert.</td>
</tr>
<tr>
<td>-1.58 My name is John.</td>
</tr>
<tr>
<td>-1.59 My name’s John.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input: How old are you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.79 I don’t know.</td>
</tr>
<tr>
<td>-1.06 I’m fine.</td>
</tr>
<tr>
<td>-1.17 I’m all right.</td>
</tr>
<tr>
<td>-1.17 I’m not sure.</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>-1.64 Twenty-five.</td>
</tr>
<tr>
<td>-1.66 Five.</td>
</tr>
<tr>
<td>-1.71 Eight.</td>
</tr>
</tbody>
</table>

Li et al. (2016)
Lack of Diversity

- Solution: mutual information criterion; response R should be predictive of user utterance U as well
Lack of Diversity

- Solution: mutual information criterion; response R should be predictive of user utterance U as well

- Standard conditional likelihood: \( \log P(R|U) \)

Li et al. (2016)
Lack of Diversity

- Solution: mutual information criterion; response R should be predictive of user utterance U as well

- Standard conditional likelihood: \( \log P(R|U) \)

- Mutual information: \( \log \frac{P(R,U)}{P(R)P(U)} = \log P(R|U) - \log P(R) \)

- \( \log P(R) \) can reflect probabilities under a language model

Li et al. (2016)
### Lack of Diversity

| message                                                                 | SEQ2SEQ                                      | MMI $P(T|S) - \lambda P(T)$ |
|------------------------------------------------------------------------|----------------------------------------------|------------------------------|
| I did not get the report from an mi-6 agent.                          | I don’t think that’s a good idea.            | You did the right thing, did you? |
| You haven’t been given an assignment in this case.                    | I don’t know what you are talking about.     | I’ve been looking all over for you. |
| I’m losing my grip.                                                   | I don’t know what you are talking about.     | I’m the only one in the world. |
| I am ready to help.                                                   | Come on, come on                             | I have something we need to talk about. |
| You programmed me to gather intelligence. That’s all I’ve ever done.  | You know that, don’t you?                   | You do have fun, don’t you? |
| I mean, we’d have to talk to him.                                      | I mean, I don’t know.                        | I mean, he’s a good guy. |
| This has nothing to do with him.                                       | You don’t know anything about him, do you?   | I know him as much.           |
| How come you never say it?                                            | I don’t know                                 | Because I don’t want to hurt you |

- **OpenSubtitles data**

---

Li et al. (2016)
Future of chatbots

- How deep can a conversation be without more semantic grounding? Basic facts aren’t even consistent...

<table>
<thead>
<tr>
<th>message</th>
<th>response</th>
</tr>
</thead>
<tbody>
<tr>
<td>In which city do you live now?</td>
<td>I live in Madrid.</td>
</tr>
<tr>
<td>In which country do you live now?</td>
<td>England, you?</td>
</tr>
</tbody>
</table>

Li et al. (2016) Persona...
Future of chatbots

‣ How deep can a conversation be without more semantic grounding? Basic facts aren’t even consistent...

‣ Can force chatbots to give consistent answers, but still probably not very interesting

<table>
<thead>
<tr>
<th>message</th>
<th>response</th>
</tr>
</thead>
<tbody>
<tr>
<td>In which city do you live now?</td>
<td>I live in Madrid.</td>
</tr>
<tr>
<td>In which country do you live now?</td>
<td>England, you?</td>
</tr>
</tbody>
</table>

Li et al. (2016) Persona...
Future of chatbots

- How deep can a conversation be without more semantic grounding? Basic facts aren’t even consistent...

- Can force chatbots to give consistent answers, but still probably not very interesting

- Xiaolce: Microsoft chatbot in Chinese, 20M users, average user interacts 60 times/month

<table>
<thead>
<tr>
<th>message</th>
<th>response</th>
</tr>
</thead>
<tbody>
<tr>
<td>In which city do you live now?</td>
<td>I live in Madrid.</td>
</tr>
<tr>
<td>In which country do you live now?</td>
<td>England, you?</td>
</tr>
</tbody>
</table>

Li et al. (2016) Persona...
Future of chatbots

- How deep can a conversation be without more semantic grounding? Basic facts aren’t even consistent...

- Can force chatbots to give consistent answers, but still probably not very interesting

- Xiaolce: Microsoft chatbot in Chinese, 20M users, average user interacts 60 times/month

- People do seem to like talking to them...?

<table>
<thead>
<tr>
<th>message</th>
<th>response</th>
</tr>
</thead>
<tbody>
<tr>
<td>In which city do you live now?</td>
<td>I live in Madrid.</td>
</tr>
<tr>
<td>In which country do you live now?</td>
<td>England, you?</td>
</tr>
</tbody>
</table>

Li et al. (2016) Persona...
Task-Oriented Dialogue
Task-Oriented Dialogue

- Question answering/search:
Task-Oriented Dialogue

- Question answering/search:
Question answering/search:

Google, what’s the most valuable American company?
Question answering/search:

Google, what’s the most valuable American company?

Apple
Task-Oriented Dialogue

- Question answering/search:
  
  Google, what’s the most valuable American company?

  Apple

  Who is its CEO?
Task-Oriented Dialogue

- Question answering/search:
  
  Google, what’s the most valuable American company?
  
  Apple
  
  Who is its CEO?
  
  Tim Cook
Task-Oriented Dialogue

- Personal assistants / API front-ends:
Task-Oriented Dialogue

- Personal assistants / API front-ends:
  
  Siri, find me a good sushi restaurant in Chelsea
Task-Oriented Dialogue

- Personal assistants / API front-ends:
  
  Siri, find me a good sushi restaurant in Chelsea

  Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google
Task-Oriented Dialogue

- Personal assistants / API front-ends:
  - Siri, find me a good sushi restaurant in Chelsea
  - Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google
  - How expensive is it?
Task-Oriented Dialogue

- Personal assistants / API front-ends:

  - Siri, find me a good sushi restaurant in Chelsea
  - Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google
  - How expensive is it?
  - Entrees are around $30 each
Task-Oriented Dialogue

- Personal assistants / API front-ends:

  - Siri, find me a good sushi restaurant in Chelsea
  - Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google
  - How expensive is it?
  - Entrees are around $30 each
  - Find me something cheaper
Task-Oriented Dialogue

- Personal assistants / API front-ends:
Hey Alexa, why isn’t my Amazon order here?
Task-Oriented Dialogue

- Personal assistants / API front-ends:

  Hey Alexa, why isn’t my Amazon order here?
  
  Let me retrieve your order. Your order was scheduled to arrive at 4pm today.
Task-Oriented Dialogue

- Personal assistants / API front-ends:

  - Hey Alexa, why isn’t my Amazon order here?
  - Let me retrieve your order. Your order was scheduled to arrive at 4pm today.
  - It never came
Hey Alexa, why isn’t my Amazon order here?

Let me retrieve your order. Your order was scheduled to arrive at 4pm today.

It never came

Okay, I can put you through to customer service.
Air Travel Information Service (ATIS)

- Given an utterance, predict a domain-specific semantic interpretation

<table>
<thead>
<tr>
<th>Utterance</th>
<th>How much is the cheapest flight from Boston to New York tomorrow morning?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal:</td>
<td>Airfare</td>
</tr>
<tr>
<td>Cost.Relative</td>
<td>cheapest</td>
</tr>
<tr>
<td>Depart.City</td>
<td>Boston</td>
</tr>
<tr>
<td>Arrival.City</td>
<td>New York</td>
</tr>
<tr>
<td>Depart.Date.Relative</td>
<td>tomorrow</td>
</tr>
<tr>
<td>Depart.Time.Period</td>
<td>morning</td>
</tr>
</tbody>
</table>

- Can formulate as semantic parsing, but simple slot-filling solutions (classifiers) work well too

DARPA (early 1990s), Figure from Tur et al. (2010)
Full Dialogue Task

- Parsing / language understanding is just one piece of a system

Young et al. (2013)
Full Dialogue Task

- Parsing / language understanding is just one piece of a system.
- Dialogue state: reflects any information about the conversation (e.g., search history).

Young et al. (2013)
Full Dialogue Task

- Parsing / language understanding is just one piece of a system
- Dialogue state: reflects any information about the conversation (e.g., search history)
- User utterance -> update dialogue state -> take action (e.g., query the restaurant database) -> say something

Young et al. (2013)
Full Dialogue Task
Full Dialogue Task

Find me a good sushi restaurant in Chelsea
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
Full Dialogue Task

Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
Full Dialogue Task

Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
curr_result <- execute_search()
Full Dialogue Task

Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
curr_result <- execute_search()

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
curr_result <- execute_search()

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea

curr_result <- execute_search()

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?

get_value(cost, curr_result)
Find me a good sushi restaurant in Chelsea

```r
restaurant_type <- sushi
current_location <- Chelsea

curr_result <- execute_search()
```

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?

```r
google_value(cost, curr_result)
```

Entrees are around $30 each
POMDP-based Dialogue Systems

- POMDP: user is the "environment," an utterance is a noisy signal of state

Young et al. (2013)
POMDP-based Dialogue Systems

- POMDP: user is the “environment,” an utterance is a noisy signal of state
- Dialogue model: can look like a parser or any kind of encoder model

Young et al. (2013)
POMDP-based Dialogue Systems

- POMDP: user is the “environment,” an utterance is a noisy signal of state
- Dialogue model: can look like a parser or any kind of encoder model
- Generator: use templates or seq2seq model

Young et al. (2013)
POMDP-based Dialogue Systems

- POMDP: user is the “environment,” an utterance is a noisy signal of state
- Dialogue model: can look like a parser or any kind of encoder model
- Generator: use templates or seq2seq model
- Where do rewards come from?

Young et al. (2013)
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea

curr_result <- execute_search()

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?
...

Okay make me a reservation!

+1 make_reservation(curr_result)
Reward for completing task?

Find me a good sushi restaurant in Chelsea

Very indirect signal of what should happen up here

restaurant_type <- sushi
location <- Chelsea
curr_result <- execute_search()

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?
...

Okay make me a reservation!

+1 make_reservation(curr_result)
Find me a good sushi restaurant in Chelsea

```
restaurant_type <- sushi
location <- Chelsea
curr_result <- execute_search()
```

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?

```
get_value(cost, curr_result)
```

+1 Entrees are around $30 each
Find me a good sushi restaurant in Chelsea

How does the user know the right search happened?

restaurant_type <- sushi
location <- Chelsea
curr_result <- execute_search()

Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google

How expensive is it?

get_value(cost, curr_result)

Entrees are around $30 each
Wizard-of-Oz

- Learning from demonstrations: “wizard” pulls the levers and makes the dialogue system update its state and take actions

Kelley (early 1980s), Ford and Smith (1982)
Full Dialogue Task

Find me a good sushi restaurant in Chelsea
Find me a good sushi restaurant in Chelsea

wizard enters these

\[
\begin{align*}
\text{restaurant}_\text{type} & \leftarrow \text{sushi} \\
\text{location} & \leftarrow \text{Chelsea} \\
\text{curr}\_\text{result} & \leftarrow \text{execute}\_\text{search}()
\end{align*}
\]
Find me a good sushi restaurant in Chelsea.

wizard enters these

\[
\begin{align*}
\text{restaurant_type} & \leftarrow \text{sushi} \\
\text{location} & \leftarrow \text{Chelsea} \\
\text{curr_result} & \leftarrow \text{execute_search}()
\end{align*}
\]

wizard types this out or invokes templates

\{ 
Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google
\}
Full Dialogue Task

Find me a good sushi restaurant in Chelsea

wizard enters these

\[
\begin{align*}
\text{restaurant_type} & \leftarrow \text{sushi} \\
\text{location} & \leftarrow \text{Chelsea} \\
\text{curr_result} & \leftarrow \text{execute_search()} \\
\end{align*}
\]

wizard types this out or invokes templates

\[
\begin{align*}
\text{Sushi Seki Chelsea is a sushi restaurant in Chelsea with 4.4 stars on Google} \\
\end{align*}
\]

- Wizard can be a trained expert and know exactly what the dialogue systems is supposed to do
Learning from Static Traces

Using either wizard-of-Oz or other annotations, can collect static traces and train from these

Bordes et al. (2017)
Full Dialogue Task

Find me a good sushi restaurant in Chelsea

```r
restaurant_type <- sushi
location <- Chelsea

curr_result <- execute_search()
```
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
stars <- 4+
curr_result <- execute_search()
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
stars <- 4+
curr_result <- execute_search()

- User asked for a “good” restaurant — does that mean we should filter by star rating? What does “good” mean?
Find me a good sushi restaurant in Chelsea

restaurant_type <- sushi
location <- Chelsea
stars <- 4+
curr_result <- execute_search()

- User asked for a “good” restaurant — does that mean we should filter by star rating? What does “good” mean?

- Hard to change system behavior if training from static traces, especially if system capabilities or desired behavior change
Goal-oriented Dialogue

- Tons of industry interest!

- Startups:
  - ASAPP
  - Maluuba
  - Eloquent Labs
  - semanticmachines
  - KITT.AI
  - VIV

- Big Companies: Apple Siri (VocalIQ), Google Allo, Amazon Alexa, Microsoft Cortana, Facebook M, Samsung Bixby, Tencent WeChat

- Lots of cool work that’s not public yet
Other Dialogue Applications
Search/QA as Dialogue

“Has Chris Pratt won an Oscar?” / “Has he won an Oscar?”
Dialogue is a very natural way to find information from a search engine or a QA system.
QA as Dialogue

- Dialogue is a very natural way to find information from a search engine or a QA system.

- Challenges:
  - QA is hard enough on its own.

Iyyer et al. (2017)
Dialogue is a very natural way to find information from a search engine or a QA system.

Challenges:
- QA is hard enough on its own
- Users move the goalposts

Iyyer et al. (2017)
QA as Dialogue

- UW QuAC dataset: Question Answering in Context

Choi et al. (2018)
Google can deal with misspellings, so more misspellings happen — Google has to do more!
Most NLP tasks

System

Data

Better model

Error analysis
Most NLP tasks

- Error analysis
- System
- Better model
- Data

- Fixed distribution (e.g., natural language sentences), error rate $\rightarrow 0$
Fixed distribution (e.g., natural language sentences), error rate $\rightarrow 0$
Dialogue Mission Creep

Most NLP tasks

- System
- Data
- Error analysis
- Better model

Fixed distribution (e.g., natural language sentences), error rate $\rightarrow 0$

Dialogue/Search/QA

- System
- Data
- Harder Data
- Better model
- Error analysis

-???
Fixed distribution (e.g., natural language sentences), error rate -> 0

Error rate -> ???; “mission creep” from HCI element
High visibility — your product has to work really well!
Takeaways

- Some decent chatbots, applications: predictive text input, ...

- Task-oriented dialogue systems are growing in scope and complexity

- More and more problems are being formulated as dialogue — interesting applications but challenging to get working well