Dialogue agents

Christopher Potts

CS 244U: Natural language understanding
May 21
Plan and goals

1. Language in context, language as action
2. The Switchboard Dialog-Act Corpus: modeling interactional linguistic patterns
3. The Cards Corpus: task-oriented dialogue
4. PLOW: dialogue agents for the Web browser
5. A framework for increasingly sophisticated dialogue agents:
   a. Markov Decision Processs (MDP)
   b. Partially Observable MDP (POMDP)
   c. Decentralized POMDP (Dec-POMDP)
   d. Approximations of b and c
Plan and goals

1. Language in context, language as action
2. The Switchboard Dialog-Act Corpus: modeling interactional linguistic patterns
3. The Cards Corpus: task-oriented dialogue
4. PLOW: dialogue agents for the Web browser
5. A framework for increasingly sophisticated dialogue agents:
   a. Markov Decision Processes
   b. Partially Observable MDP
   c. Decentralized POMDP
   d. Approximations of b and c

Deeply influenced by Adam Vogel, who successfully defended his Stanford PhD thesis this morning!
Language is action

Winograd (1986:170):

“all language use can be thought of as a way of activating procedures within the hearer. We can think of an utterance as a program – one that indirectly causes a set of operations to be carried out within the hearer’s cognitive system.”
Eliza (Weizenbaum 1966)

1. User: You are like my father in some ways.
2. ELIZA: What resemblance do you see
3. User: You are not very aggressive but I think you don’t want me to notice that.
4. ELIZA: What makes you think I am not aggressive?
5. User: You don’t argue with me.
6. ELIZA: Why do you think I don’t argue with you?
7. User: You are afraid of me.
8. ELIZA: Does it please you to believe I am afraid of you?

Rewrite user’s reply by (i) swapping 1st and 3rd person, (ii) interpolating stock phrases, and (iii) using scores to rank possible transformations.
SHRDLU (Winograd 1972)

Full-grounded system that parses the user’s input, maps it to logical form, interprets that logical form in its world, tries to take appropriate action, and generates responses.

http://hci.stanford.edu/winograd/shrdlu/

One project did succeed. Terry Winograd’s program SHRDLU could use English intelligently, but there was a catch: the only subject you could discuss was a micro-world of simulated blocks.
SHRDLU (Winograd 1972)

1. Person: WHAT DOES THE BOX CONTAIN?
2. Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
3. Person: WHAT IS THE PYRAMID SUPPORTED BY?
4. Computer: THE BOX.
5. Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
6. Computer: FOUR OF THEM.
7. Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
9. Person: IS IT SUPPORTED?
10. Computer: YES, BY THE TABLE.
• In the 1967 Stanley Kubrick movie *2001: A Space Odyssey*, the spaceship’s computer HAL can
  • display graphics;
  • play chess; and
  • conduct natural, open-domain conversations with humans.

• How well did the filmmakers do at predicting what computers would be capable in 2001?

(Slide idea from Andrew McCallum)
HAL

Jurassic Park (1993)

(Slide idea from Andrew McCallum)
Overview & motivations
SwDA
PLOW
MDPs & grounded semantics
The Cards Corpus
POMDPs & approximate Dec-POMDPs
Refs.

HAL

Chess

Deep Blue (1997)

(Slide idea from Andrew McCallum)
HAL

Dialogue

HAL

David Bowman: Open the pod bay doors, HAL.
HAL: I’m sorry, Dave, I’m afraid I can’t do that.
David: What are you talking about, HAL?
HAL: I know that you and Frank were planning to disconnect me, and I’m afraid that’s something I cannot allow to happen.

2014

(Slide idea from Andrew McCallum)
You: Any good burger joints around here?
Siri: I found a number of burger restaurants near you.
You: Hmm. How about tacos?
Apple: [Siri remembers that you asked about restaurants. so it will look for Mexican restaurants in the neighborhood. And Siri is proactive, so it will question you until it finds what you’re looking for.]
Siri

Colbert: For the love of God, the cameras are on, give me something?

Siri: What kind of place are you looking for? Camera stores or churches?

[...]  

Colbert: I don’t want to search for anything! I want to write the show!

Siri: Searching the Web for “search for anything. I want to write the shuffle.”

(Slide from Marie de Marneffe)
Language in context

Intended functional meaning

Get hot coffee
Get B to get me hot coffee
Convince B that this coffee is cold

Interpreted functional meaning

Get A hot coffee
A wants hot coffee
Believe that this coffee is cold

“This coffee is cold”

Referential meaning

yum, coffee!
A’s connotative meaning

yuck, coffee!
B’s connotative meaning

Speaker (A) Interpreter (B)
Discourse models

Discourse is structured by an evolving set of abstract, implicit, issues about which the participants have only partial knowledge:

**Questions under Discussion (QUDs)**
McCarthy 1980; Groenendijk and Stokhof 1984; Rooth 1985; Lewis 1988; Ginzburg 1996a; Roberts 1996; Büring 1999

**Decision problems (games)**
Lewis 1969; Clark 1996; Merin 1997; Blutner 1998; Parikh 2001; Beaver 2002; van Rooy 2003; Benz et al. 2005; Franke 2009

**Goal-orientation**

For much more: [http://www.ling.ohio-state.edu/~croberts/QUDbib/](http://www.ling.ohio-state.edu/~croberts/QUDbib/)
Pragmatically required over-answering

**Context:** Homer calls a hotel.

Homer: Is Lisa Simpson in Room 10?
Clerk A: She’s in room 20.
Clerk B: #No.

Which room is Lisa in?
- Is Lisa in 10?
- Is Lisa in 20?
- Is Lisa in 30?

(Roberts 1996; Ginzburg 1996a; Champollion 2008)
Domain restriction

I didn’t see any.

(Roberts 1996; Ginzburg 1996a; Malamud 2006)
Domain restriction

• Are there typos in my slides?

I didn’t see any.

(Roberts 1996; Ginzburg 1996a; Malamud 2006)
Domain restriction

- Are there typos in my slides?
- Are there bookstores downtown?

_I didn’t see any._

(Roberts 1996; Ginzburg 1996a; Malamud 2006)
Domain restriction

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the cupboard?

*I didn’t see any.*

(Roberts 1996; Ginzburg 1996a; Malamud 2006)
Domain restriction

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the cupboard?
- . . .

I didn’t see any.

(Roberts 1996; Ginzburg 1996a; Malamud 2006)
Granularity

*Where are you from?*

(Groenendijk and Stokhof 1984; Ginzburg 1996b)
Granularity

*Where are you from?*

- *Connecticut.* (Issue: birthplaces)

(Groenendijk and Stokhof 1984; Ginzburg 1996b)
Granularity

Where are you from?

- Connecticut. (Issue: birthplaces)
- The U.S. (Issue: nationalities)

(Groenendijk and Stokhof 1984; Ginzburg 1996b)
Granularity

Where are you from?

- Connecticut. (Issue: birthplaces)
- The U.S. (Issue: nationalities)
- Stanford. (Issue: affiliations)

(Groenendijk and Stokhof 1984; Ginzburg 1996b)
Granularity

Where are you from?

- Connecticut. (Issue: birthplaces)
- The U.S. (Issue: nationalities)
- Stanford. (Issue: affiliations)
- Planet earth. (Issue: intergalactic meetings)

(Groenendijk and Stokhof 1984; Ginzburg 1996b)
Summary of corpus resources

- **SwDA:**

- **SwDA with Treebank3 alignment:**
  [http://compprag.christopherpotts.net/swda.html](http://compprag.christopherpotts.net/swda.html)

- **Edinburgh Map Corpus:**
  [http://groups.inf.ed.ac.uk/maptask/](http://groups.inf.ed.ac.uk/maptask/)

- **TRIPS:**

- **TRAINS:**

- **Cards:**
  [http://CardsCorpus.christopherpotts.net/](http://CardsCorpus.christopherpotts.net/)

- **SCARE:**
  [http://slate.cse.ohio-state.edu/quake-corpora/scare/](http://slate.cse.ohio-state.edu/quake-corpora/scare/)

- **The Carnegie Mellon Communicator Corpus (human–computer):**
  [http://www.speech.cs.cmu.edu/Communicator/](http://www.speech.cs.cmu.edu/Communicator/)
A decision-theoretic framework for dialogue agents

Figure: MDP

Figure: POMDP

Figure: Dec-POMDP
The Switchboard Dialog-Act Corpus

- The SwDA extends the Switchboard-1 Telephone Speech Corpus, Release 2, with turn/utterance-level dialog-act tags.
- The tags summarize syntactic, semantic, and pragmatic information about the associated turn.
- It is freely available:
  
  http://www.stanford.edu/~jurafsky/ws97/

- The SwDA is not inherently linked to the Penn Treebank 3 parses of Switchboard, and it is far from straightforward to align the two resources (Calhoun et al. 2010).
- In addition, the SwDA is not distributed with the Switchboard’s tables of metadata about the conversations and their participants.
- I created a CSV version of the corpus that pools all of this information to the best of my ability, thereby allowing study of the correlations among dialog tags, conversational metadata, and full syntactic structures:

  http://compprag.christopherpotts.net/swda.html
Example dialogue

^h A.1 utt1: {F Uh, } let’s see. /
% A.1 utt2: How [ about, + {F uh, } let’s see, about ] ten
qo A.1 utt3: {F uh, } what do you think was different ten ye
sv B.2 utt1: {D Well, } I would say as, far as social change
sv B.2 utt2: [ They, + they ] did more things together. /
 b @A.3 utt1: Uh-huh <> . /
sv B.4 utt1: {F Uh, } they ate dinner at the table together.
sv B.4 utt2: {F Uh, } the parents usually took out [ time, +
b A.5 utt1: Uh-huh. /
sv B.6 utt1: {F Uh, } although I’m not a mother, [ I, + I ]
qo B.6 utt2: {F Uh, } what # do you # --
% A.7 utt1: # We, # - /
+ B.8 utt1: -- think about that? /

Table: FILENAME: 4360_1599_1589
There are over 200 tags in the SwDA, most used only a few times. It is more common to work with a collapsed version involving just 44 tags.

<table>
<thead>
<tr>
<th>name</th>
<th>act</th>
<th>tag</th>
<th>example</th>
<th>train count</th>
<th>full count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement-non-opinion</td>
<td>sd</td>
<td>sd</td>
<td>Me, I'm in the legal department.</td>
<td>72824</td>
<td>75145</td>
</tr>
<tr>
<td>Acknowledge (Backchannel)</td>
<td>b</td>
<td>b</td>
<td>Uh-huh.</td>
<td>37096</td>
<td>38298</td>
</tr>
<tr>
<td>Statement-opinion</td>
<td>sv</td>
<td>sv</td>
<td>I think it's great</td>
<td>25197</td>
<td>26428</td>
</tr>
<tr>
<td>Agree/Accept</td>
<td>aa</td>
<td>aa</td>
<td>That's exactly it.</td>
<td>10820</td>
<td>11133</td>
</tr>
<tr>
<td>Abandoned or Turn-Exit</td>
<td>%</td>
<td>%</td>
<td>So, -</td>
<td>10569</td>
<td>15550</td>
</tr>
<tr>
<td>Appreciation</td>
<td>ba</td>
<td>ba</td>
<td>I can imagine.</td>
<td>4633</td>
<td>4765</td>
</tr>
<tr>
<td>Yes-No-Question</td>
<td>qy</td>
<td>qy</td>
<td>Do you have to have any special training?</td>
<td>4624</td>
<td>4727</td>
</tr>
<tr>
<td>Non-verbal</td>
<td>x</td>
<td>x</td>
<td>[Laughter], [Throat.clearing]</td>
<td>3548</td>
<td>3630</td>
</tr>
<tr>
<td>Yes answers</td>
<td>ny</td>
<td>ny</td>
<td>Yes.</td>
<td>2934</td>
<td>3034</td>
</tr>
<tr>
<td>Conventional-closing</td>
<td>fc</td>
<td>fc</td>
<td>Well, it's been nice talking to you.</td>
<td>2486</td>
<td>2582</td>
</tr>
<tr>
<td>Uninterpretable</td>
<td>%</td>
<td>%</td>
<td>But, uh, yeah</td>
<td>2158</td>
<td>3034</td>
</tr>
<tr>
<td>Wh-Question</td>
<td>qw</td>
<td>qw</td>
<td>Well, how old are you?</td>
<td>1911</td>
<td>1979</td>
</tr>
<tr>
<td>No answers</td>
<td>nn</td>
<td>nn</td>
<td>No.</td>
<td>1340</td>
<td>1377</td>
</tr>
<tr>
<td>Response Acknowledgement</td>
<td>bk</td>
<td>bk</td>
<td>Oh, okay.</td>
<td>1277</td>
<td>1306</td>
</tr>
<tr>
<td>Hedge</td>
<td>h</td>
<td>h</td>
<td>I don't know if I'm making any sense or not.</td>
<td>1182</td>
<td>1226</td>
</tr>
<tr>
<td>Declarative Yes-No-Question</td>
<td>qyd</td>
<td>qyd</td>
<td>So you can afford to get a house?</td>
<td>1174</td>
<td>1219</td>
</tr>
<tr>
<td>Other fo_o_fw_by_bc</td>
<td></td>
<td></td>
<td>Well give me a break, you know.</td>
<td>1074</td>
<td>883</td>
</tr>
<tr>
<td>Backchannel in question form</td>
<td>bh</td>
<td>bh</td>
<td>Is that right?</td>
<td>1019</td>
<td>1053</td>
</tr>
<tr>
<td>Quotation</td>
<td><code>q</code></td>
<td><code>q</code></td>
<td>You can't be pregnant and have cats.</td>
<td>934</td>
<td>983</td>
</tr>
<tr>
<td>Summarize/reformulate</td>
<td>bf</td>
<td>bf</td>
<td>Oh, you mean you switched schools for the kids.</td>
<td>919</td>
<td>952</td>
</tr>
<tr>
<td>Affirmative non-yes answers</td>
<td>na</td>
<td>na</td>
<td>It is.</td>
<td>836</td>
<td>847</td>
</tr>
<tr>
<td>Action-directive</td>
<td>ad</td>
<td>ad</td>
<td>Why don’t you go first</td>
<td>719</td>
<td>746</td>
</tr>
</tbody>
</table>
DAMSL tags

There are over 200 tags in the SwDA, most used only a few times. It is more common to work with a collapsed version involving just 44 tags.

<table>
<thead>
<tr>
<th>name</th>
<th>act</th>
<th>tag</th>
<th>train example</th>
<th>full example</th>
<th>count</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 Collaborative Completion</td>
<td>^2</td>
<td></td>
<td>Who aren’t contributing.</td>
<td></td>
<td>699</td>
<td>723</td>
</tr>
<tr>
<td>24 Repeat-phrase</td>
<td>b^m</td>
<td></td>
<td>Oh, fajitas</td>
<td></td>
<td>660</td>
<td>688</td>
</tr>
<tr>
<td>25 Open-Question</td>
<td>qo</td>
<td></td>
<td>How about you?</td>
<td></td>
<td>632</td>
<td>656</td>
</tr>
<tr>
<td>26 Rhetorical-Questions</td>
<td>qh</td>
<td></td>
<td>Who would steal a newspaper?</td>
<td></td>
<td>557</td>
<td>575</td>
</tr>
<tr>
<td>27 Hold before answer/agreement</td>
<td>^h</td>
<td></td>
<td>I’m drawing a blank.</td>
<td></td>
<td>540</td>
<td>556</td>
</tr>
<tr>
<td>28 Reject</td>
<td>ar</td>
<td></td>
<td>Well, no</td>
<td></td>
<td>338</td>
<td>346</td>
</tr>
<tr>
<td>29 Negative non-no answers</td>
<td>ng</td>
<td></td>
<td>Uh, not a whole lot.</td>
<td></td>
<td>292</td>
<td>302</td>
</tr>
<tr>
<td>30 Signal-non-understanding</td>
<td>br</td>
<td></td>
<td>Excuse me?</td>
<td></td>
<td>288</td>
<td>298</td>
</tr>
<tr>
<td>31 Other answers</td>
<td>no</td>
<td></td>
<td>I don’t know</td>
<td></td>
<td>279</td>
<td>286</td>
</tr>
<tr>
<td>32 Conventional-opening</td>
<td>fp</td>
<td></td>
<td>How are you?</td>
<td></td>
<td>220</td>
<td>225</td>
</tr>
<tr>
<td>33 Or-Clause</td>
<td>qrr</td>
<td></td>
<td>or is it more of a company?</td>
<td></td>
<td>207</td>
<td>209</td>
</tr>
<tr>
<td>34 Dispreferred answers</td>
<td>arp_nd</td>
<td></td>
<td>Well, not so much that.</td>
<td></td>
<td>205</td>
<td>207</td>
</tr>
<tr>
<td>35 3rd-party-talk</td>
<td>t3</td>
<td></td>
<td>My goodness, Diane, get down from there.</td>
<td></td>
<td>115</td>
<td>117</td>
</tr>
<tr>
<td>36 Offers, Options, Commits</td>
<td>oo_co_cc</td>
<td></td>
<td>I’ll have to check that out</td>
<td></td>
<td>109</td>
<td>110</td>
</tr>
<tr>
<td>37 Self-talk</td>
<td>t1</td>
<td></td>
<td>What’s the word I’m looking for</td>
<td></td>
<td>102</td>
<td>103</td>
</tr>
<tr>
<td>38 Downplayer</td>
<td>bd</td>
<td></td>
<td>That’s all right.</td>
<td></td>
<td>100</td>
<td>103</td>
</tr>
<tr>
<td>39 Maybe/Accept-part</td>
<td>aap_am</td>
<td></td>
<td>Something like that</td>
<td></td>
<td>98</td>
<td>105</td>
</tr>
<tr>
<td>40 Tag-Question</td>
<td>^g</td>
<td></td>
<td>Right?</td>
<td></td>
<td>93</td>
<td>92</td>
</tr>
<tr>
<td>41 Declarative Wh-Question</td>
<td>qw^d</td>
<td></td>
<td>You are what kind of buff?</td>
<td></td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>42 Apology</td>
<td>fa</td>
<td></td>
<td>I’m sorry.</td>
<td></td>
<td>76</td>
<td>79</td>
</tr>
<tr>
<td>43 Thanking</td>
<td>ft</td>
<td></td>
<td>Hey thanks a lot</td>
<td></td>
<td>67</td>
<td>78</td>
</tr>
</tbody>
</table>
Switchboard Dialog Act Corpus with parsetrees

- My release of the SwDA includes the Treebank3 POS tags.

- It also includes the Treebank3 trees, but these are somewhat more challenging to work with:
  - Only 118,218 (53%) of utterances have trees.
  - The Treebank3 team merged some utterances into single trees.
  - Other utterances were split across trees.
  - The turn numbering was altered, often dramatically.

- On the bright side:
  - 82% of the utterances with trees correspond to a single tree.
  - With the exception of non-verbal (x) and tag-questions (ˆg), the distribution of tags in this subset is basically the same as the distribution for the whole corpus.

- Additional details: http://compprag.christopherpotts.net/swda.html
Act-tag and syntactic category alignment
A quick experiment: to what extent are dialog act tags and clause-types aligned?

1 Request act
   a. Take these pills twice a day.
   b. You should take these twice a day.
   c. Could you please take these twice a day?

2 Question act
   a. Is today Tuesday?
   b. It’s Tuesday, right?
   c. I need to confirm that it’s Tuesday.

3 Imperative form
   a. Take these pills twice a day.
   b. Have a seat.
   c. Get well soon.

4 Interrogative
   a. Is today Tuesday?
   b. Is he ever tall!
   c. Can you pass the salt?
Act-tag and syntactic category alignment

A quick experiment: to what extent are dialog act tags and clause-types aligned?

The hearer’s perspective: given that I heard a syntactic structure with root label $L$, what are the speaker’s possible intended dialog acts?
Act-tag and syntactic category alignment

A quick experiment: to what extent are dialog act tags and clause-types aligned?

The speaker’s perspective: given that I want to convey dialog act $D$, what is the best structure for me to choose?
Modeling act sequences

- Modeling act sequences could be an important step towards realistic interpretation and production.
- Shriberg et al. (1998) and Stolcke et al. (2000) use acoustic features to predict general dialog act labels, using the SwDA. Their model is a decision-tree classifier.
- Other classifiers might also be appropriate; the natural assumption here is that the classifications decisions are made on a by-utterance basis, with no inspection of neighboring utterances (Bangalore et al. 2006; Kumar Rangarajan Sridhar et al. 2009).
- Dialog act prediction can also be viewed as a sequence modeling problem akin to POS tagging, and thus Hidden Markov Models and Conditional Random Fields models are often used. Such models incorporate earlier and/or later tags to make classification decisions.
On the SwDA for dialogue research

Advantages

- Richly annotated.
- Includes speech data.
- Includes sociolinguistic metadata.
- Long conversations, and lots of them.
- Participants did not typically know each other before the conversation, so most of their common ground is general knowledge.
On the SwDA for dialogue research

Advantages

- Richly annotated.
- Includes speech data.
- Includes sociolinguistic metadata.
- Long conversations, and lots of them.
- Participants did not typically know each other before the conversation, so most of their common ground is general knowledge.

Disadvantages

- Open-domain, unfocussed (participants do not stick to their topics).
- Virtually no hope of modeling the context or grounding the language in the world or in action.
PLOW: webpage structure as context

For the PLOW system, the context is the webpage:

- From NL Interpretation:
  (FILL :object *FULLNAME*)

- From GUI Interpretation:
  (action-performed
   (FILL-FIELD
    :object INPUT22
    :value “Harold Pinter”))

System learns the rule:
- Whenever we have to find an object in which we want to type an instance of some concept,
- Look for an input node that has a sibling node of type text whose content is some linguistic realization of that concept

Figure 4: Learning to find and fill a text field

- Project homepage:
  http://www.cs.rochester.edu/research/cisd/projects/plow/

- Language processing with the TRIPS parser:
  http://www.cs.rochester.edu/research/cisd/projects/trips/parser/cgi/web-parser-xml.cgi
Learning new rules and generalizations

- Learning rules of the form ‘If A, then B, else C’ is a challenge because the latent variable A is generally not observed. Rather, one sees only B or C.
- In an interactive, instructional setting, one needn’t rely entirely on abduction or probabilistic inference: users generally state the needed rules during their interactions.
Language-based principles

1. The user’s actions ground the parsed language.
   - Put the name here. (user clicks on the DOM element)
   - This is the ISBN number. (user highlights some text)
   - Find another tab. (user has selected a tab)

2. The DOM structure grounds the user’s indexicals and other referential devices.
   - Put the zip code here
   - Put hotels here

3. Indefinites mark new information; definites refer to established information:
   - A man walked in. He/The man looked tired.
   - an address ⇒ new input parameter
   - the address ⇒ existing input parameter
Interaction and error correction

- PLOW is tested with human users in real scenarios. (It has been used by the US Military Health System to set up doctor’s appointments.)
- Thus, PLOW tries to immediately apply the rules it infers, so that the user will correct it. This helps with:
  - finding the right level of generalization; and
  - overcoming noise in the context (from poor HTML mark-up)
Evaluation

16 independent evaluators trained on PLOW and three other systems.

**Phase 1**

1. The evaluators taught the systems some predefined tasks.
2. The system then performed those tasks with different input parameters.

**Phase 2**

1. The evaluators used the systems to teach some of the tasks at right.
2. PLOW received the highest average score of all systems.
3. Evaluators had free choice of which system to use. 13 chose PLOW for at least one task, and PLOW was chosen for 30 of the 55 evaluation tasks.

---

1. What <businesses> are within <distance> of <address>?
2. Get directions for <integer> number of restaurants within <distance> of <address>.
3. Find articles related to <topic> written for <project>.
4. Which <project> had the greatest travel expenses between <start date> and <end date>?
5. What is the most expensive purchase approved between <start date> and <end date>?
6. For what reason did <person> travel for <project> between <start date> and <end date>?
7. Find <ground-transport, parking> information for <airport>.
8. Who should have been notified that <person> was out of the office between <start date> and <end date>?
9. Summarize all travel and purchase costs for <project> between <date> and <date> by expense category.
10. Which projects exceeded the current government maximum allowable expense for travel costs?

*Figure 1: Previously unseen tasks used in the evaluation*
Markov Decision Processes (MDPs)

- The agent has complete knowledge of the environment and its own current state, but the effects of its actions are non-deterministic.

- MDPs were developed starting in the 1950s by Richard Bellman (1957), Ronald Howard (1960), Karl Åström (1965), Edward Sondik (1971), Richard Sutton (1988), and others. Most of this work concerns efficiently finding the agent's optimal action.

- Howard (1978) describes one of the earliest applications: programming the Sears, Roebuck, and Co.’s giant Addressograph mechanical computer to optimize the process of choosing which customers to send which catalogues (late 1950s): “The optimum policy was confirmed by applying it to [...] a selected set of customers whose purchases were very carefully monitored. When the policy was later implemented on the full customer set, the results closely confirmed the model predictions” (p. 100).
Defined

Definition (MDP)

1. $S$ is a finite set of states.
2. $A$ is a finite set of actions.
3. $R : (S \times A) \mapsto \mathbb{R}$ is the reward function.
4. $T : (S \times A \times S) \mapsto [0, 1]$ is the state transition function.

Example

Cab driver Ron serves towns A and B. He has two actions: cruise for fares or wait at a cab stand.

<table>
<thead>
<tr>
<th>cruise</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>B</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>stand</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>B</td>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

(a) $T$ for cruising around

(b) $T$ for the cab stand

Table: Optimizing Ron’s plans based on his data.

(c) $R$
Optimization

Definition (Bellman operator for MDPs)

Define $B_0(s) = 0$ for all $s \in S$. Then for all $t > 0$:

$$B_t(s, a) = R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') B_{t-1}(s')$$

where $0 < \gamma \leq 1$ is a discounting term (a dollar today is worth more than a dollar tomorrow).

**Value Iteration** ($S, A, R, T, \gamma, \varepsilon$)

1. $V(s) = 0, V'(s) = 0$ for all $s \in S$
2. while True
3. for $s \in S$
   # argmax for policy too:
4. $V'(s) = \max_{a \in A} [R(s, a) + \gamma \sum_{s' \in S} T(s, a, s') V(s')]$
5. if $|V'(s) - V(s)| < \varepsilon$ for all $s \in S$
   return $V'$
6. else $V = V'$
Optimal planning under uncertainty

Example

Cab driver Ron serves towns A and B. He has two actions: cruise for fares or wait at a cab stand.

\[
\begin{array}{ccc}
\text{cruise} & A & B \\
A & 0.9 & 0.1 \\
B & 0.1 & 0.9 \\
\end{array}
\quad \begin{array}{ccc}
\text{stand} & A & B \\
A & 0.4 & 0.6 \\
B & 0.6 & 0.4 \\
\end{array}
\]

(a) $T$ for cruising around
(b) $T$ for the cab stand

\[
\begin{array}{ccc}
cruise & A & B \\
\text{stand} & $8$ & $20$ \\
\text{stand} & $5$ & $22$ \\
\end{array}
\]

(c) $R$

A $\leftrightarrow$ stand \quad B $\leftrightarrow$ cruise

(d) Optimal policy

Table: Optimizing Ron's plans based on his data.
A simple robot controller (Russell and Norvig 2003:§17)

Figure: Action-specific state transitions

(a) Optimal policy when the reward (penalty) for being in a blank square is $-0.04$.

(b) Optimal policy when the reward (penalty) for being in a blank square is $-0.3$.

Figure: Optimality for different reward functions.
Vogel and Jurafsky (2010)

- Agents that learn to follow naviational instructions on maps.
- MDP formulation with online reinforcement learning.
- **Inspiring idea**: feature functions $\phi(s, a)$ and associated learned weights, to process unknown utterances, landmarks, etc.
- **Inspiring idea**: learning probabilistic word meanings from the interaction of language, the world, and the rewards.
- Limitations begin to show us the need for more complex agents.
The Edinburgh Map Corpus (Thompson et al. 1993)

One participant tells the other how to reproduce a path through a map.

g  right it starts directly above the crest falls if you go
to the left of your page just to the edge of the crest
falls
f  mmhmm

g  come south due south to the bottom of the page
f  mmhmm

g  go to the left of the page to about an inch from the
end
f  over the banana tree

g  i suppose so yeah eh
f  mmhmm

g  go north to the level of the footbridge
f  mmhmm

g  go up and go across the footbridge and stop
exactl— right at the end edge of the footbridge
f  above the footbridge

g  o— over the footbridge
f  mm

g  and stop right at the end of it

g  there is a poisoned stream on mine but which you
don’t have

Transcripts, audio, maps, etc.: [http://groups.inf.ed.ac.uk/maptask/](http://groups.inf.ed.ac.uk/maptask/)
MDP formulation and learning

1. $S$: a set of $s = (u, l, c)$ triples:
   - A set of utterances $u$
   - A set of landmarks $l$
   - $c \in \{\text{North, South, East, West}\}$

2. $A$: $(l, c)$, meaning pass $l$ on side $c$

3. $R((u, l, c), (l', c')) = \begin{bmatrix} \mathbb{1}[l = l'] + \\ \mathbb{1}[c = c'] + \\ \text{sim}(u, l') \end{bmatrix}$

4. $T(s, a) = s'$

5. $\phi(s, a) \in \mathbb{R}^n$ capturing world and linguistic information
MDP formulation and learning

1. **S**: a set of $s = (u, l, c)$ triples:
   - A set of utterances $u$
   - A set of landmarks $l$
   - $c \in \{\text{North, South, East, West}\}$

2. **A**: $(l, c)$, meaning pass $l$ on side $c$

3. $R\left((u, l, c), (l', c')\right) = \begin{bmatrix} \mathbb{I}[l = l'] & + \\ \mathbb{I}[c = c'] & + \\ \text{sim}(u, l') & \end{bmatrix}$

4. $T(s, a) = s'$

5. $\phi(s, a) \in \mathbb{R}^n$ capturing world and linguistic information

---

Input: Dialog set $D$
- Reward function $R$
- Feature function $\phi$
- Transition function $T$
- Learning rate $\alpha_t$

Output: Feature weights $\theta$
1. Initialize $\theta$ to small random values
2. until $\theta$ converges do
3. foreach Dialog $d \in D$ do
4. Initialize $s_0 = (l_1, u_1, \theta)$,
   $a_0 \sim \Pr(a_0|s_0; \theta)$
5. for $t = 0; s_t$ non-terminal; $t++$ do
6. Act: $s_{t+1} = T(s_t, a_t)$
7. Decide: $a_{t+1} \sim \Pr(a_{t+1}|s_{t+1}; \theta)$
8. Update:
   \[
   \Delta \leftarrow R(s_t, a_t) + \theta^T\phi(s_{t+1}, a_{t+1}) \\
   - \theta^T\phi(s_t, a_t)
   \]
   $\theta \leftarrow \theta + \alpha_t \phi(s_t, a_t)\Delta$
9. end
10. end
11. return $\theta$

Algorithm 1: The SARSA learning algorithm.
Learned paths

Figure 4: Sample output from the SARSA policy. The dashed black line is the reference path and the solid red line is the path the system follows.
Learned meanings

Figure 5: This figure shows the relative weights of spatial features organized by spatial word. The top row shows the weights of allocentric (landmark-centered) features. For example, the top left figure shows that when the word *above* occurs, our policy prefers to go to the north of the target landmark. The bottom row shows the weights of egocentric (absolute) spatial features. The bottom left figure shows that given the word *above*, our policy prefers to move in a southerly cardinal direction.
The Cards Corpus

http://CardsCorpus.christopherpotts.net/

Included

- The transcripts in CSV format
- Python classes for working with the transcripts
- Examples of the Python classes in action
- R code for reading in the corpus as a data frame
- All the annotations used in the work described here

By the numbers

- 1,266 transcripts
- Game length mean: 373.21 actions (median 305, sd 215.20)
- Card pickup: 19,157
- Card drop: 12,325
- Move: 371,811
- Utterance: 45,805 (260,788 words, ≈4,000 word vocab)
Amazon Mechanical Turk HIT (Human Intelligence Task)

- **Title:** Collaborative Search Game with Chat
- **Description:** Two-player collaborative video game involving dialogue/chat with other Turkers.
- **Payment:** $1.00, and up to $0.50 cents for rich, collaborative problem-solving using meaningful dialogue.
- **Restrictions:** US IP addresses; at least 95% approval rating
- **Timing:** mid-week, 7:00 am – 3:00 pm Pacific time
- **Turker Nation:** posting on Turker Nation about our HIT and its goals, responding to Turkers’ questions and concerns, and learning from Turkers about what life is like for them.
Amazon Mechanical Turk HIT (Human Intelligence Task)

**Instructions**

1. You'll first be presented with an experimental consent form; the transcripts from these games will be used to study collaborative problem solving.
2. Once you click through, you'll go into "Wait for Participant" mode until another player shows up. It might take a while for another player to show up! We suggest opening a new window and doing other things until someone arrives.
3. When play begins, you'll be presented with a task description: you and your partner are tasked with finding six consecutive cards of the same suit.
4. When you've solved the game, click the "Task Complete" button to register that the task is complete. You can continue playing and chatting at that point, or you can click **Complete HIT and return to MTurk** at the top of the screen.
5. You will receive a bonus of up to $0.50 for good collaboration with meaningful communication.
6. Notes:
   - Once the game begins, you can move around with the arrow keys or the on-screen buttons.
   - The chat window is near the top of the screen.
   - You can't see your partner, and your partner can't see you!
   - The yellow boxes mark cards in your "line of sight".
7. Caution: Players who do not complete the task properly will not be paid. It is vital that you read and understand the task description and follow the instructions.
8. Feel free to play multiple times — especially useful if you get good at the game.

**Annotated game screen**
Amazon Mechanical Turk HIT (Human Intelligence Task)

TYPE HERE

Yellow boxes mark cards in your line of sight.

You are on 2D

Task description: Six consecutive cards of the same suit

The cards you are holding

Move with the arrow keys or these buttons.
Amazon Mechanical Turk HIT (Human Intelligence Task)

Gather six consecutive cards of a particular suit (decide which suit together), or determine that this is impossible. Each of you can hold only three cards at a time, so you’ll have to coordinate your efforts. You can talk all you want, but you can make only a limited number of moves.
Amazon Mechanical Turk HIT (Human Intelligence Task)

Gather six consecutive cards of a particular suit (decide which suit together), or determine that this is impossible. Each of you can hold only three cards at a time, so you’ll have to coordinate your efforts. You can talk all you want, but you can make only a limited number of moves.

What’s going on?
⇒
Which suit should we pursue?
⇒
Which sequence should we pursue?
⇒
Where is card X?
## Transcripts: environment metadata

<table>
<thead>
<tr>
<th>Agent</th>
<th>Time</th>
<th>Action type</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server</td>
<td>0</td>
<td>COLLECTION_SITE</td>
<td>Amazon Mechanical Turk</td>
</tr>
<tr>
<td>Server</td>
<td>0</td>
<td>TASK_COMPLETED</td>
<td>2010-06-17 10:10:53 EDT</td>
</tr>
<tr>
<td>Server</td>
<td>0</td>
<td>PLAYER_1</td>
<td>A00048</td>
</tr>
<tr>
<td>Server</td>
<td>0</td>
<td>PLAYER_2</td>
<td>A00069</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>P1_MAX_LINEOFSIGHT</td>
<td>3</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>P2_MAX_LINEOFSIGHT</td>
<td>3</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>P1_MAX_CARDS</td>
<td>3</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>P2_MAX_CARDS</td>
<td>3</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>P1_MAX_TURNS</td>
<td>200</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>P2_MAX_TURNS</td>
<td>200</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>GOAL_DESCRIPTION</td>
<td>Gather six consecutive cards ...</td>
</tr>
<tr>
<td>Server</td>
<td>2</td>
<td>CREATE_ENVIRONMENT</td>
<td>[ASCII representation]</td>
</tr>
<tr>
<td>Player 1</td>
<td>2092</td>
<td>PLAYER_INITIALIZER</td>
<td>16,15</td>
</tr>
<tr>
<td>Player 2</td>
<td>2732</td>
<td>PLAYER_INITIALIZER</td>
<td>9,10</td>
</tr>
</tbody>
</table>
Transcripts: environment metadata

NEW_SECTION

1,2:2D;1,7:KH;1,7:9S;1,11:6C;1,13:QC;1,14:QS;
2,18:3H;2,18:9H;
3,19:4H;4,8:AC;4,19:3D;
4,19:KD;
5,14:QH;5,15:5S;5,15:2S;5,16:4D;5,16:10C;5,18:4S;
6,11:KC;6,15:9C;
7,11:2H;7,13:7S;
8,2:QD;8,4:AD;8,11:JC;8,20:8S;
9,9:10S;9,9:6H;9,9:8C;9,10:7H;9,14:JS;
10,1:2C;10,10:8D;11,14:6D;11,14:10H;
11,18:4C;11,18:9D;
12,10:3S;12,12:6S;12,16:5H;12,16:JD;12,20:3C;
13,4:5C;13,4:JH;13,15:KS;
14,2:5D;14,20:10D;15,2:AH;

---
## Transcripts: game play

<table>
<thead>
<tr>
<th>Agent</th>
<th>Time</th>
<th>Action type</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 1</td>
<td>566650</td>
<td>PLAYER_MOVE</td>
<td>7,11</td>
</tr>
<tr>
<td>Player 2</td>
<td>567771</td>
<td>CHAT_MESSAGE_PREFIX</td>
<td>which c’s do you have again?</td>
</tr>
<tr>
<td>Player 1</td>
<td>576500</td>
<td>CHAT_MESSAGE_PREFIX</td>
<td>i have a 5c and an 8c</td>
</tr>
<tr>
<td>Player 2</td>
<td>577907</td>
<td>CHAT_MESSAGE_PREFIX</td>
<td>i jsut found a 4 of clubs</td>
</tr>
<tr>
<td>Player 1</td>
<td>581474</td>
<td>PLAYER_PICKUP_CARD</td>
<td>7,11:8C</td>
</tr>
<tr>
<td>Player 1</td>
<td>586098</td>
<td>PLAYER_MOVE</td>
<td>7,10</td>
</tr>
</tbody>
</table>
Novice strategy

Player 1: Hello. Are you here?
Player 2: yes
Player 2: do you see any cards
Player 1: Yes. I see a yellow spot. Those are our cards. We'll only be able to see the ones that are in our view
Player 1: until we move with our arrows.
Player 2: i see 3 of them
Player 1: We only have a certain number of moves, so we should decide how we’re going to do this before we use them, do you think?
Player 2: sure
Player 1: Ok. So, we have to pick up six cards of the same suit, in a row...
Player 1: each of us can hold three, so...
Player 1: I think I should get my three, then you should get your three or vice versa
Player 2: ok
Player 2: you go ahead
Player 1: What suit should we do?
Player 1: And which six cards do you want to try for?
Player 2: whatever you want
Player 1: I'm Courtney, by the way- nice to meet you.
Player 2: i'm becky....nice to meet you too
Player 1: Hi Becky. How about we go for hearts? And take 234567

[...]

I'm assuming that the conversation is set in a card game scenario where the players are trying to strategize how to pick up cards of the same suit in a row. The players are discussing their options and making decisions based on the information they have and the rules of the game.
Journeyman strategy

These players have explored and are now forming a strategy:

Player 1  I have 9 clubs and K clubs
Player 1  want to look for clubs?
Player 2  ok

[...]

The players then find various clubs, checking with each other frequently, until they gain an implicit understanding of which specific sequences to try for (either 8C-KC or 9C-AC):

Player 1  so you are holding Jc and Kc now?
Player 2  i now have 10d JC and KC
Player 2  yes
Player 1  drop 10d and look for either 8c or Ace of clubs
Expert strategy

Player 2: hi
Player 1: hi--which side r u on?
Player 2: right side
Player 2: u?
Player 1: left/middle
Player 1: ok i gathered everything in my area
Player 2: i think i have all of them also
Player 1: how bout 5C - 10C?
Player 2: ok
Player 1: i have 5C, 8C, 9C, and you should have 6C, 7C, 10C
Player 2: got them
Asymmetric play

Player 1: very limited number of moves but infinite line-of-sight; Player 2: large number of moves but very limited line of sight.

Player 1: Hi
Player 2: hi where are you
Player 1: near the upper right
Player 2: ok any cards that way
Player 1: lots of cards near me to the upper right corner
Player 2: did you get that
Player 1: get wjat ?
Player 2: the drop in the top right
Player 1: I have not gone there yet
Player 2: ok I’ll wait
Player 2: we have the 4 8 j h
Player 2: 3 k c
Player 1: ok
Player 1: the cards are pretty scattered
Player 1: did you check the entire right column?

...
Language in context

Each transcript is a data structure that is intuitively a list of temporally-ordered states

\[
\text{(context, event)}
\]

The context includes

- local information (the state of play at that point)
- historical information (the events up to that point)
- global information (limitations of the game, the task, etc.)

When the event is an utterance, we can interpret it *in context*.

This is what pragmatics is all about, but it is very rare to have a dataset that truly lets you do it.
Task-oriented dialogue corpora

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Task type</th>
<th>Domain</th>
<th>Task-orient.</th>
<th>Docs.</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switchboard</td>
<td>discussion</td>
<td>open</td>
<td>very loose</td>
<td>2,400</td>
<td>aud/txt</td>
</tr>
<tr>
<td>SCARE</td>
<td>search</td>
<td>3d world</td>
<td>tight</td>
<td>15</td>
<td>aud/vid/txt</td>
</tr>
<tr>
<td>TRAINS</td>
<td>routes</td>
<td>map</td>
<td>tight</td>
<td>120</td>
<td>aud/txt</td>
</tr>
<tr>
<td>Map Task</td>
<td>routes</td>
<td>map</td>
<td>tight</td>
<td>128</td>
<td>aud/vid/txt</td>
</tr>
<tr>
<td>Columbia Games</td>
<td>games</td>
<td>maps</td>
<td>tight</td>
<td>12</td>
<td>aud/txt</td>
</tr>
<tr>
<td>Cards</td>
<td>search</td>
<td>2d grid</td>
<td>tight</td>
<td>1,266</td>
<td>txt in context</td>
</tr>
</tbody>
</table>

Chief selling points for Cards:

- Pretty large.
- Controlled enough that similar things happen often.
- Very highly structured — the only corpus whose release version allows the user to replay all games with perfect fidelity.
Papers using the Cards corpus

- Djalali et al. (2012): anapora and domain restriction
- Djalali et al. (2011): presuppositions
- Potts (2012): goal-orientation of underspecified locative expressions
- Vogel et al. (2013a): emergent Gricean behavior with Dec-POMDPs
- Vogel et al. (2013b): conversational implicature with Dec-POMDPs
POMDPs and approximate Dec-POMDPs

We want our agent to:

- Make moves that are likely to lead it to the card.
- Change its behavior based on observations it receives.
- Respond to locative advice from the other player.
- Give locative advice to the other player.

Modeling the problem as a POMDP allows us to train agents that have these properties.
Simplified cards scenario

Both players must find the ace of spades.

DialogBot:
Grounded language interpretation

“in the bottom you see the opening on the bottom row”
\[ \Downarrow \]
BOARD(entrance & bottom); \( H: 5.48 \)

“in the top right of the middle part of the board”
\[ \Downarrow \]
middle(top & right); \( H: 5.27 \)

“i’m in the center”
\[ \Downarrow \]
BOARD(middle); \( H: 7.37 \)

Utterances as bags of words. No preprocessing (yet) for spelling correction, lemmatization, etc. Assign semantic tags using log-linear classifiers trained on the corpus data.
POMDPs

The agent has only probabilistic information about its current state (and the effects of its actions are non-deterministic, as in MDPs).

**Definition (POMDP)**

A POMDP is a structure \((S, A, R, T, \Omega, O)\):

- \((S, A, R, T)\) is an MDP.
- \(\Omega\) is a finite set of observations.
- \(O : (A \times S \times \Omega) \mapsto [0, 1]\) is the observation function.
ListenerBot (a POMDP agent)

- **S**: all combinations of the player’s region and the card’s region
- **$b_0$**: initial belief state (distribution over $S$)
- **A**: *travel* actions for each region, and a single *search* action
- **$\Omega$**: \{AS seen, AS not seen\}
- **$\Sigma$**: a set of messages, treated as observations; each message $\sigma$ denotes a distribution $P(s \mid \sigma)$ over states $s$. We apply Bayes rule to incorporate these into the POMDP observations.
- **T**: distributions $P(s' \mid s, a)$, except *travel* actions fail between nonadjacent regions
- **O**: distributions $P(o \mid s, a)$; *travel* actions never return positive observations; *search* actions return positive observations only if the player’s current region contains the AS
- **R**: small negative for not being on the card, large positive for being on it. No sensitivity to the other player.
Optimization

A belief state for \((S, A, R, T, \Omega, O)\) is a probability distribution \(b\) over \(S\).

\[
P(s, a, o, b) = O(s, a, o) \sum_{s' \in S} T(s', a, s) b(s')
\]  

(1)

\[
b_o^a(s) = \frac{P(s, a, o, b)}{\sum_{s' \in S} P(s', a, o, b)}
\]  

(2)

**Definition (Bellman operator for POMDPs)**

Let \(b\) be a belief state for \((S, A, R, T, \Omega, O)\). Set \(P_0(b') = 0\) for all belief states \(b'\). Then for all \(t > 0\):

\[
P_t(b, a) = \left( \sum_{s \in S} b(s) R(s, a) \right) + \gamma \sum_{o \in \Omega} \left( \sum_{s \in S} P(s, a, o, b) \right) P_{t-1}(b_o^a)
\]

where \(0 < \gamma \leq 1\) is a discounting term.
Approximate solutions take us (only) part of the way

- An exact solution specifies the value of every action at any reachable belief state.

- In practice, only approximate solutions are tractable. We used the PERSEUS solution algorithm.

- Even approximate solutions are generally only possible for problems with $< 10K$ states.

<table>
<thead>
<tr>
<th>Card location</th>
<th>Agent location</th>
<th>Partner location</th>
<th>Partner’s card beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>231 $\times$ 231</td>
<td>231 $\times$ 231</td>
<td>231 $\times$ 231</td>
<td>231</td>
</tr>
<tr>
<td>$\approx 50K$</td>
<td>$\approx 12M$</td>
<td>$\approx 3B$</td>
<td></td>
</tr>
</tbody>
</table>

Table: Size of the state-space for the one-card game.
Language as a representation for planning

- Divide the board up into $n$ regions, for some tractable $n$
- Generate this partition using our locative phrase distributions.
- $k$-means clustering in locative phrase space.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>.01</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.04</th>
<th>.12</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOARD(bbottom;entrance):</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>0</td>
</tr>
<tr>
<td>BOARD(middle;top):</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>.07</td>
<td>.08</td>
<td>.09</td>
<td>.10</td>
<td>.02</td>
<td>0</td>
</tr>
<tr>
<td>BOARD(bottom;corner;right):</td>
<td>0</td>
<td>0</td>
<td>.11</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.06</td>
<td>0</td>
</tr>
</tbody>
</table>

...
Clusters induced

Figure: 12-cell clustering.
Clusters induced

Figure: 14-cell clustering.
Clusters induced

Figure: 16-cell clustering.
Clusters induced

Figure: 18-cell clustering.
ListenerBot example

ListenerBot:
ListenerBot example

ListenerBot:
### ListenerBot example

**ListenerBot:**

```
- "it's on the left side"
```

![Diagram of a maze with a rabbit and a card]
ListenerBot example

ListenerBot:

“it’s on the left side”

BOAR(D(left)
ListenerBot example

ListenerBot:
DialogBot (an approximate Dec-POMDP)

DialogBot is a strict extension of ListenerBot:

- The set of states is now all combinations of
  - both players’ positions
  - the card’s region
  - the region the other player believes the card to be in
- The set of actions now includes dialog actions.
- (The player assumes that) a dialog action $U$ alters the other player’s beliefs in the same way that $U$ would impact his own beliefs.
- Same basic reward structure as for Listenerbot, except now also sensitive to whether the other player has found the card.
Belief-state approximation

(a) Exact multi-agent belief tracking

(b) Approximate multi-agent belief tracking
How the agents relate to each other

Figure: In the full Dec-POMDP (b), both agents receive individual observations and choose actions independently. Optimal decision making requires tracking all possible histories of beliefs of the other agent. DialogBot approximates the full Dec-POMDP as single-agent POMDP. At each time step, DialogBot marginalizes out the possible observations $\bar{o}$ that ListenerBot received, yielding an expected belief state $\bar{b}$. 

(a) ListenerBot POMDP
(b) Full Dec-POMDP
(c) DialogBot POMDP
DialogBot and ListenerBot play together

**DialogBot beliefs:**
ListenerBot’s position

**ListenerBot beliefs:**
ListenerBot’s beliefs
DialogBot and ListenerBot play together

DialogBot beliefs: ListenerBot’s position

ListenerBot beliefs: ListenerBot’s beliefs
DialogBot and ListenerBot play together

DialogBot: “Top”

DialogBot beliefs

ListenerBot beliefs

DialogBot beliefs: ListenerBot’s position

DialogBot beliefs: ListenerBot’s beliefs
DialogBot and ListenerBot play together

DialogBot beliefs: ListenerBot’s position

DialogBot beliefs: ListenerBot’s beliefs

ListenerBot beliefs

Dialogbot: “Top”
DialogBot and ListenerBot play together

### DialogBot beliefs

- **ListenerBot’s position**

### ListenerBot beliefs

- **DialogBot’s beliefs**
DialogBot and ListenerBot play together

DialogBot beliefs

ListenerBot beliefs

DialogBot beliefs: ListenerBot’s position

DialogBot beliefs: ListenerBot’s beliefs
DialogBot and ListenerBot play together

DialogBot beliefs: ListenerBot’s position

ListenerBot beliefs: ListenerBot’s beliefs
Grown-up DialogBots (a week of policy exploration)
Baby DialogBots (a few hours of policy exploration)
## Experimental results

<table>
<thead>
<tr>
<th>Agents</th>
<th>Success</th>
<th>Average Moves</th>
</tr>
</thead>
<tbody>
<tr>
<td>ListenerBot &amp; ListenerBot</td>
<td>84.4%</td>
<td>19.8</td>
</tr>
<tr>
<td>ListenerBot &amp; DialogBot</td>
<td>87.2%</td>
<td>17.5</td>
</tr>
<tr>
<td>DialogBot &amp; DialogBot</td>
<td>90.6%</td>
<td>16.6</td>
</tr>
</tbody>
</table>

**Table**: The evaluation for each combination of agents. 500 random initial states per agent combination. It pays to model other minds!
Emergent pragmatics

**Quality**

- The Gricean maxim of quality says roughly “Be truthful”.
- For DialogBot, this emerges from the decision problem: false information is (typically) more costly.
- DialogBot would lie if he thought it would move them toward the objective.

**Quantity and Relevance**

- The Gricean maxims of quantity and relevance for informative, timely contributions.
- When DialogBot finds the card, he communicates the information, not because he is hard-coded to do so, but rather because it will help the other player find it.
References I


Calhoun, Sasha; Jean Carletta; Jason Brenier; Neil Mayo; Daniel Jurafsky; Mark Steedman; and David I. Beaver. 2010. The NXT-format Switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation* 44(4):387–419.


Djalali, Alex; Sven Lauer; and Christopher Potts. 2012. Corpus evidence for preference-driven interpretation. In Maria Aloni; Vadim Kimmelman; Floris Roelofsen; Galit Weidman Sassoon; Katrin Schulz; and Matthijs Westera, eds., *Proceedings of the 18th Amsterdam Colloquium: Revised Selected Papers*, 150–159. Berlin: Springer.


References II


Merin, Arthur. 1997. If all our arguments had to be conclusive, there would be few of them. Arbeitspapiere SFB 340 101, University of Stuttgart, Stuttgart.


Shriberg, Elizabeth; Rebecca Bates; Paul Taylor; Andreas Stolcke; Daniel Jurafsky; Klaus Ries; Noah Coccaro; Rachel Martin; Marie Meteer; and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? Language and Speech 41(3–4):439–487.


Stolcke, Andreas; Klaus Ries; Noah Coccaro; Elizabeth Shriberg; Rebecca Bates; Daniel Jurafsky; Paul Taylor; Rachel Martin; Marie Meteer; and Carol Van Ess-Dykema. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics 26(3):339–371.
Stone, Matthew; Richmond Thomason; and David DeVault. 2007. Enlightened update: A computational architecture for presupposition and other pragmatic phenomena. To appear in Donna K. Byron; Craige Roberts; and Scott Schwenter, Presupposition Accommodation.


