Snippets of Research

George Konidaris
gdk@cs.brown.edu

Fall 2017
“If we knew what we were doing, it wouldn’t be called research.”
Research

*The production of new knowledge.*

Largely carried out by:

- University research labs
  - Professors (PIs), postdocs, graduate students
- Federal Labs (e.g., Los Alamos)
- Industrial research labs (e.g., Google DeepMind)

Results subsequently published (i.e., typically given away).

Largely funded by:

- The federal government (why?)
  - NSF, NIH, DARPA, etc.
- Some industry funding.
Communication is About Hidden State

A human trying to communicate something is like an HMM.

• True “meaning” not observed.
• Gesture observed
• Speech observed
• What they “mean” may change over time.
Models

(a) Uninformed Transitions (no dependency on corpus or previous states).

(b) Unigram model (dependency on corpus, but not previous states).

(Whitney et al., 2016)
(c) Bigram model (dependency on corpus as well as one previous state).
(d) Trigram model (dependency on corpus as well as two previous states).

(Whitney et al., 2016)
Please hand me a bowl

(Whitney et al., 2016)
### Results

**TABLE I**

**Simulated Context, Language, and Gesture**

(a) Results using Gesture without Language

<table>
<thead>
<tr>
<th>Model</th>
<th>d = 3</th>
<th>d = 5</th>
<th>d = 10</th>
<th>d = ∞</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>23.41% ± 1.73%</td>
<td>15.49% ± 1.46%</td>
<td>8.84% ± 1.15%</td>
<td>0.67% ± 0.329%</td>
</tr>
<tr>
<td>Unigram</td>
<td>34.82% ± 1.94%</td>
<td>27.74% ± 1.83%</td>
<td>19.21% ± 1.60%</td>
<td>5.43% ± 0.92%</td>
</tr>
<tr>
<td>Bigram</td>
<td>42.74% ± 2.01%</td>
<td>35.73% ± 1.94%</td>
<td>28.23% ± 1.83%</td>
<td>12.68% ± 1.34%</td>
</tr>
<tr>
<td>Trigram</td>
<td>41.04% ± 1.99%</td>
<td>32.50% ± 1.91%</td>
<td>27.38% ± 1.81%</td>
<td>12.74% ± 1.35%</td>
</tr>
</tbody>
</table>

(b) Results Using Gesture with Ambiguous Language Requests

<table>
<thead>
<tr>
<th>Model</th>
<th>d = 3</th>
<th>d = 5</th>
<th>d = 10</th>
<th>d = ∞</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>74.39% ± 1.78%</td>
<td>70.91% ± 1.84%</td>
<td>67.13% ± 1.91%</td>
<td>47.99% ± 2.03%</td>
</tr>
<tr>
<td>Unigram</td>
<td>75.61% ± 1.74%</td>
<td>72.56% ± 1.81%</td>
<td>70.61% ± 1.84%</td>
<td>52.74% ± 2.03%</td>
</tr>
<tr>
<td>Bigram</td>
<td>77.80% ± 1.69%</td>
<td>76.22% ± 1.73%</td>
<td>72.56% ± 1.81%</td>
<td>53.11% ± 2.03%</td>
</tr>
<tr>
<td>Trigram</td>
<td>77.38% ± 1.69%</td>
<td>75.12% ± 1.76%</td>
<td>72.68% ± 1.81%</td>
<td>53.72% ± 2.03%</td>
</tr>
</tbody>
</table>

(c) Results Using Gesture with Unambiguous Language Requests

<table>
<thead>
<tr>
<th>Model</th>
<th>d = 3</th>
<th>d = 5</th>
<th>d = 10</th>
<th>d = ∞</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>94.63% ± 0.92%</td>
<td>93.96% ± 0.97%</td>
<td>93.41% ± 1.00%</td>
<td>87.50% ± 1.35%</td>
</tr>
<tr>
<td>Unigram</td>
<td>95.12% ± 0.87%</td>
<td>94.27% ± 0.94%</td>
<td>94.39% ± 0.94%</td>
<td>89.09% ± 1.27%</td>
</tr>
<tr>
<td>Bigram</td>
<td>95.67% ± 0.82%</td>
<td>95.00% ± 0.89%</td>
<td>94.27% ± 0.94%</td>
<td>88.66% ± 1.28%</td>
</tr>
<tr>
<td>Trigram</td>
<td>95.55% ± 0.84%</td>
<td>94.70% ± 0.90%</td>
<td>94.39% ± 0.94%</td>
<td>88.41% ± 1.30%</td>
</tr>
</tbody>
</table>

(Whitney et al., 2016)
Interpreting Multimodal Referring Expressions in Real Time

Miles Eldon, David Whitney, Stefanie Tellex
Humans to Robots Laboratory
Like an HMM but ... 

You can also take actions!

Hidden state MDP
- This is called a POMDP
- Partially observable MDP

Can go further and construct a model where the robot chooses to ask a question, or not, to disambiguate.

(Whitney et al., 2017)
The Intelligent Robot Lab
Skill Hierarchies

**Hierarchical RL**: base hierarchical control on *skills*.
- Component of behavior.
- Performs continuous, low-level control.
- Can treat as discrete action.

*Behavior is modular and compositional.*

Skills are like *subroutines*.

```python
def abs(x):
    if(x > 0):
        return x
    else:
        return -x
```

[Wilkes, Wheeler and Gill, 1951]
Hierarchical RL
Hierarchical RL
The Options Framework

An option is one formal model of a skill.

An option $o$ is a policy unit:

- **Initiation set** $I_o : S \rightarrow \{0, 1\}$
- **Termination condition** $\beta_o : S \rightarrow [0, 1]$
- **Option policy** $\pi_o : S \times A \rightarrow [0, 1]$

[Sutton, Precup and Singh 1999]
Skill Acquisition

- A robot learning to solve a task
- Extracting skills from solution
- Deploying them in a new task
Training Room

Episode 1 (35x)
Acquired Skills
The Test Room
The Test Room

Median Test Performance Comparison

Without Acquired Skills  With Acquired Skills
The Test Room

[AAAI 2011]
Symbolic Planning

Problem difficulty shouldn’t depend on low-level state space.
Learning Symbolic Representations
Symbolic Planning
Learning Symbolic Representations
Symbolic Representations

(:action nav_to_cooler1
 :parameters ()
 :precondition (and (symbol0))
 :effect (and (symbol0) (not (symbol1))
 (decrease (reward) 37.25))
)
Symbolic Representations

(:action cupboard_open1
 :parameters ()
 :precondition (and (symbol1) (symbol3) (symbol4))
 :effect (and (symbol5) (not (symbol4))
         (decrease (reward) 67.44))
)

symbol1

symbol3

symbol4

symbol5
Symbolic Representations

(:action pick_up1
 :parameters ()
 :precondition (and (symbol0) (symbol8)
                      (symbol12))
 :effect (and (symbol11) (symbol2)
            (not (symbol3)) (not (symbol12))
            (decrease (reward) 52.62))
)
Symbolic Representations

(:action pick_up2
 :parameters ()
 :precondition (and (symbol1) (symbol3)
               (symbol5) (symbol6) (symbol11))
 :effect (probabilistic
  0.0559 (and)
  0.9441 (and (symbol2) (not (symbol3))
              (decrease (reward) 53.42))
 )
 )

symbol1

symbol5 and symbol6

symbol3

symbol2
Symbolic Planning
Michael Littman
MDPs
POMDPs

The definition of a state:

- Sufficient statistic of past history,
- For predicting $s'$ and $r$

$$T(s_{t+1}|s_t, a_t, s_{t-1}, a_{t-1}, ..., s_0) = T(s_{t+1}|s_t, a_t)$$
$$R(s_{t+1}, s_t, a_t, s_{t-1}, a_{t-1}, ..., s_0) = R(s_{t+1}, s_t, a_t)$$

That is what the state means.

**Very strong assumption:** the agent has access to state.
Markov and Robots

Does the robot see everything it needs to be able to predict the effects of its own actions?
Example
POMDPs

Partially observable Markov decision processes:
- Formalism for the non-Markov case
- Decision making under state uncertainty
- State uncertainty is unavoidable in real life
- The central theoretical objects for robotics

(Kaelbling, Littman, Cassandra, 1998)
POMDPs

General idea:

- *There is an MDP.*
- Agent does not observe state directly
- Instead, observations!
- Observations probabilistically generated from state.
POMDPs

More formally, a POMDP is:

- \( S \), a set of states
- \( A \), a set of actions
- \( T \), transition function
- \( R \), reward function
- \( \gamma \), discount factor
- \( \Omega \), set of observations
- \( O \), observation function \( O(\omega_t|s_t) \)


\[ \text{MDP} \]
A robot is a device that induces a POMDP.
Communicating Reward Functions

Where do rewards come from?

They’re supposed to express user preferences.
Approach
Communicating Reward Functions

(Ho et al., 2015)
Communicating Reward Functions

How teachers might teach

(Feedback Function)

$F_1$

$F_2$

What each learner model would learn

($\pi_{\text{Reward-Maximizing}}$)

($\pi_{\text{Action-Feedback}}$)

(Ho et al., 2015)
Communicating Reward Functions

Figure 2: An example three room floor layout with one robot and two chairs. The object reference for the red, green, and blue room are $r_1$, $r_2$, and $r_3$, respectively. The references for the yellow and blue chair are $c_1$ and $c_2$, respectively.

(Peng et al., 2015)
Training

(a) Train 1-2  (b) Train 3  (c) Train 4  (d) Train 5  (e) Train 6  (f) Train 7

<table>
<thead>
<tr>
<th></th>
<th>Command Given</th>
<th>Associated Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>go to the green room</td>
<td>4a</td>
</tr>
<tr>
<td>2.</td>
<td>go to the red room</td>
<td>4a</td>
</tr>
<tr>
<td>3.</td>
<td>go to the blue room</td>
<td>4b</td>
</tr>
<tr>
<td>4.</td>
<td>go to the yellow room</td>
<td>4c</td>
</tr>
<tr>
<td>5.</td>
<td>take the blue chair to the yellow room</td>
<td>4d</td>
</tr>
<tr>
<td>6.</td>
<td>take the purple bag to the blue room</td>
<td>4e</td>
</tr>
<tr>
<td>7.</td>
<td>take the yellow bag to the red room</td>
<td>4f</td>
</tr>
</tbody>
</table>

(Peng et al., 2015)
**Testing**

<table>
<thead>
<tr>
<th>#</th>
<th>Test commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>take the purple chair to the green room</td>
</tr>
<tr>
<td>2</td>
<td>take the blue bag to the yellow room</td>
</tr>
<tr>
<td>3</td>
<td>go to the green room</td>
</tr>
<tr>
<td>4</td>
<td>go to the red room</td>
</tr>
<tr>
<td>5</td>
<td>go to the blue room</td>
</tr>
</tbody>
</table>

(Peng et al., 2015)
Faculty

Regular Faculty

R. Iris Bahar
Professor of Engineering and Computer Science
Profile • Home Page

Theophilus Benson
Assistant Professor of Computer Science
Profile

Ugur Cetintemel
Professor of Computer Science and Department Chair
Database Systems, Distributed Systems, Data Science
Profile • Home Page

Eugene Charniak
University Professor of Computer Science
Deep Learning, Artificial Intelligence, Machine Learning
Profile • Home Page

Tom Doepchner
Associate Professor (Research) of Computer Science and Vice Chair
Computer Systems
Profile • Home Page

Pedro Felipe Felzenszwalb
Professor of Engineering and Computer Science
Artificial Intelligence, Machine Learning, Algorithms and Theory
Profile • Home Page

Kathi Fisler
Professor (Research) of Computer Science and Associate Director of
Profile "n"

Rodrigo Fonseca
Associate Professor of Computer Science
Dartmouth, 1956