Robots Learning (Through) Interactions

May 7th 2020

Jens Kober
Advanced Robot Hardware + Manual Programming

https://youtu.be/fRj34o4hN4I
What Makes Tasks Hard?

- Complex dynamics
- Uncertainties and variations
  - Objects
  - Environment
  - Tasks
  - Human behaviour
- Occurring in all robot domains!

Learning to Interact
Why is Robotics Different?

IBM Research

sociable.co

europe1.fr

DeepMind

openclipart.org
(Reinforcement) Learning in Robotics

- Safe with real robots
- Fast learning
  - Sample efficient – “small” data
  - Incorporate prior knowledge
  - Few open parameters
- Real time computations

• Domain appropriate methods needed!
Naïve Deep RL

https://youtu.be/vv85S4Z-ZG0
Imitation Learning
• Teacher demonstrates skill, student tries to mimic

Reinforcement Learning
• Practice, practice, practice…
How to Learn?

• Continued student-teacher interaction
  – Additional demonstrations
  – Intermittent feedback

• Largely missing in robot learning!

• Benefits
  – Speed-up
  – Complex tasks
  – Intuitive

Interacting to Learn
Learning to Interact
Imitation Learning

- Combinatorial explosion
- Force → Position? Position → Force?

![Diagram showing demonstration and generalization processes.](image)
Kober, Gienger, & Steil, ICRA 2015

https://youtu.be/t_ZoiKcEM0M
Primitives within a Sequence

Movement Primitive 1

Movement Primitive 2

Movement Primitive 3

Sensor Data

Kinesthetic Demonstrations

Kober, Gienger, & Steil, ICRA 2015
Primitives within a Sequence

Score per MP/frame/component/modality based on statistics & contact

converging/attractor  constant  diverging

Scores: Movement Primitive 2

<table>
<thead>
<tr>
<th></th>
<th>position</th>
<th>force</th>
</tr>
</thead>
<tbody>
<tr>
<td>world</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>90.0</td>
<td>20.0</td>
</tr>
<tr>
<td>z</td>
<td>10.0</td>
<td>16.0</td>
</tr>
<tr>
<td>object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>8.0</td>
<td>20.0</td>
</tr>
<tr>
<td>z</td>
<td>10.0</td>
<td>16.0</td>
</tr>
</tbody>
</table>
Objective: maximize expected return $J(\theta)$
Reward-Weighted Imitation

- Maximize reward = optimize strategy
- One possibility: reward-weighted imitation

\[
\theta' = \theta + E \left\{ \sum_{t=1}^{T} Q_t^{sa} W_t^{s} \right\}^{-1} E \left\{ \sum_{t=1}^{T} Q_t^{sa} W_t^{s} \varepsilon_t \right\}
\]

Kober & Peters, NIPS 2008
Reward-weighted Imitation

Trial 1
Return 0.1

Trial 2
Return 0.3

Trial 3
Return 0.8

Trial 4
Return 1.0
https://youtu.be/cNyoMVZQdYM
Reinforcement Learning:
- Learn behaviors using trial and error and a scalar reward function

State Representation Learning:
- Cost functions acting on the state representation
- Prior knowledge about the world
- Force learning a more general representation
• Auto encoding
• Instantaneous reward prediction
• (Inverse) state dynamics
• Slowness and diversity
• Reinforcement Learning
Experiment

(optional) Pre-training

Training / testing tracks
Train Track

RL only

RL + SRL
Splitting SRL and RL

Phase 1
DQN/DDPG
freeze

Phase 2
CMA-ES

\( \tilde{s} = g(o; \theta_s) \)

\( a = h(\tilde{s}; \theta_a) \)

\( \hat{y}_{srl_1}(\tilde{s}, \cdot; \theta_{srl_1}) \)
\( \vdots \)
\( \hat{y}_{srl_n}(\tilde{s}, \cdot; \theta_{srl_n}) \)

\( \hat{Q}(\tilde{s}, a; \theta_Q) \)
rotate box 180 degrees counterclockwise
Questions so far?
Reinforcement Learning: Exploration & Human Advice

Sutton & Barto 1998

Model human advice as exploration
https://youtu.be/pto8NZdum2s
Back to Imitation Learning...
Interactive Learning: COACH

\[ \pi(s_t; \theta) \]

Teacher

Action \( a_t \)

State \( s_t \)

Correction \( h_t \)

Environment

Pérez, Celemin Paez, Ruiz-del-Solar, & Kober, ICRA 2019
Deep COACH - Offline

Encoder \[\rightarrow\] Decoder

Encoder \[\rightarrow\] Decoder

Policy
Deep COACH - Online

Encoder

Decoder

Policy

$$S_t^{HD}$$

$$a_t$$

$$S_t$$

$$\hat{S}_t^{HD}$$

Pérez, Celemin Paez, Ruiz-del-Solar, & Kober, ICRA 2019
https://youtu.be/i4f1D4CH26E
Interactive Learning

Pérez, Celemin Paez, Ruiz-del-Solar, & Kober, ICRA 2019

https://youtu.be/i4f1D4CH26E
Deep COACH - Memory

Transition Model

Encoder \( o_t \) \( h_{t-1} \) \( h_t \) LSTM \( a_t \) \( o_t \)

Decoder \( \tilde{o}_{t+1} \)

Policy

Carlos Celemin

Pérez, Celemin Paez, Franzese, Ruiz-del-Solar, & Kober, RAM 2020
Model Learning using D-COACH and SRL
Teaching Robots: A Child's Play

Walt Disney Studios – Real Steel

https://youtu.be/apxeeW-hm6I
Conclusion
Summary

• Robot learning constraints
  – “Small” data
  – Safety

• Interactive Learning
  – Efficient & intuitive

• Challenges
  – Uncertainty
  – Variations
Questions?

j.kober@tudelft.nl

www.jenskober.de