Rewards in reinforcement learning for robot assisted therapy

Vilde Gjærum and Paul Maroldt

Summer 2018
Introduction on Autism Spectrum Disorder

- Problems in social behavior
- Repetitive patterns of behavior
Introduction on behavioral therapy

Goals:
- reinforcing typical behavior
- suppress atypical behavior
- diminish symptoms
Introduction on Applied Behavior Analysis

- Discrete Trial Learning
- Incidental Teaching
- Pivotal Response Training
Reinforcement Learning and Rewards

1) task

2) action

therapist

patient

3) reward
Research questions

- Which are the rewards that should be used in therapy of individuals with ASD?
Delmonte et al. (2012): Social and monetary reward processing in autism spectrum disorders

from: Delmonte et al. (2012)
Delmonte et al. (2012): Social and monetary reward processing in autism spectrum disorders

from: Delmonte et al. (2012)
Delmonte et al. (2012): Social and monetary reward processing in autism spectrum disorders

Activation of left dorsal striatum:

<table>
<thead>
<tr>
<th></th>
<th>ASD Group</th>
<th>Test Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>monetary</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>social</td>
<td>low</td>
<td>high</td>
</tr>
</tbody>
</table>

from: Delmonte et al. (2012)
Kohls et al. (2011): Atypical Brain Responses to Reward Cues in Autism as Revealed by Event-Related Potentials
Kohls et al. (2011): Atypical Brain Responses to Reward Cues in Autism as Revealed by Event-Related Potentials

<table>
<thead>
<tr>
<th></th>
<th>ASD</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Go hit rate (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-reward</td>
<td>81,3 (13,8)</td>
<td>87,8 (10,9)</td>
</tr>
<tr>
<td>Social</td>
<td>87,2 (11,2)</td>
<td>92,7 (6,2)</td>
</tr>
<tr>
<td>Monetary</td>
<td>88,3 (10,0)</td>
<td>94,4 (5,5)</td>
</tr>
<tr>
<td><strong>No-go rejection rate (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-reward</td>
<td>94,9 (5,4)</td>
<td>96,4 (7,4)</td>
</tr>
<tr>
<td>Social</td>
<td>95,2 (7,8)</td>
<td>97,9 (3,3)</td>
</tr>
<tr>
<td>Monetary</td>
<td>99,1 (2,6)</td>
<td>99,5 (1,5)</td>
</tr>
</tbody>
</table>

from: Kohls et al. (2011)
Solomon et al. (2015): Feedback-Driven Trial-by-Trial Learning in Autism Spectrum Disorders

from: Solomon et al. (2015)
Solomon et al. (2015): Feedback-Driven Trial-by-Trial Learning in Autism Spectrum Disorders

from: Solomon et al. (2015)
Solomon et al. (2015): Feedback-Driven Trial-by-Trial Learning in Autism Spectrum Disorders

<table>
<thead>
<tr>
<th></th>
<th>ASD Group</th>
<th>TD Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>win-stay</td>
<td>0.73 (0.16)</td>
<td>0.82 (0.11)</td>
</tr>
<tr>
<td>lose-shift</td>
<td>0.35 (0.16)</td>
<td>0.33 (0.13)</td>
</tr>
</tbody>
</table>

from: Solomon et al. (2015)
## Comparison

<table>
<thead>
<tr>
<th></th>
<th>ASD vs. TD</th>
<th>Monetary vs. Social in ASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delmonte et al.</td>
<td>-</td>
<td>Less brain response to social rewards</td>
</tr>
<tr>
<td>Kohls et al.</td>
<td>General reward processing deficit</td>
<td>Just slight differences</td>
</tr>
<tr>
<td>Scott-Van Zeeland et al.</td>
<td>General learning deficit</td>
<td>Less brain response to social rewards</td>
</tr>
<tr>
<td>Schuetz et al.</td>
<td>Disconnection between desire and rewards</td>
<td>Deficite in social reward processing</td>
</tr>
<tr>
<td>Solomon et al.</td>
<td>Learning impairment</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusion on first research question

• generally dysfunctional reward system
• impairment in learning
• difficulty developing reward-based working memory
• neural difference in processing social rewards

• This may be a consequence of
  • not having learned them in early age,
  • a general impairment,
  • both.

→ many repetitions necessary
→ No general reward type
→ highly individualized in therapy
Robot assisted therapy

1) task

2) action

therapist

patient

3) reward
Reinforcement learning in machine learning

- Supervised learning vs. reinforcement learning
  RL does not need directly labeled data
  Supervised learning never better than its supervisor

- Some parameters in RL-system:
  - Policy-function
  - Action
  - State
  - Reward/Value
  - Greedy vs long-term
  - Exploration vs exploitation

- How can an agent choose what reinforces to give?
- What are the reinforcers the agent should use to update its policy?
- How does the reinforcement learning closed loop system look, i.e how do the agent and the user interact?
Adaptive robot assisted therapy using interactive reinforcement learning

- The interactive reinforcement learning agent never stops learning

<table>
<thead>
<tr>
<th>DURATION</th>
<th>EASY</th>
<th>MEDIUM</th>
<th>NORMAL</th>
<th>HARD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

from: Tsiakas et al. (2016)
Personalized machine learning for robot perception of affect and engagement in autism therapy

• Automated analysis of children's behavioral cues
  Valence - happy, sad, scared, indifferent and so on
  Arousal - is the kid paying attention to the robot?
  Engagement - is the kid attempting to solve the tasks given by the robot?

• How are these cues measured?
  Vocal language - measured by microphones
  Body/facial language - measured by cameras
  Physiological reactions (pulse, electrodermal activity and temperature) - measured by a wristband
Machine Learning in Robot Assisted Therapy (RAT)

from: Rudovic et al. (2018)
• Policy-function:
  What task/reward to choose given state of the kid

• Action:
  The set of tasks/reward to choose from

• State:
  Valence, engagement, arousal, finished task or not

• Reward:
  Valence, engagement and arousal scores, score for finishing task

• Greedy vs long-term
  How should the long-term reward be defined?

• Exploration vs exploitation
  How often should one change actions?
Robotics and Embedded Systems • Department of Computer Science • Technical University of Munich

Vilde Gjærum and Paul Maroldt

Machine Learning in Robot Assisted Therapy (RAT)

Arousal → Audio recordings
Engagement → Video recordings
Valence → Physiological recordings

Accomplishing task or not

Wrist band MUSE headset
POLICY UPDATING

1. Give task
2. Reaction to task
3. Improving tasks
4. Give reward
5. Learning
6. Reaction to reward
7. Improving rewards

ACTIONS

STATES
1. Give task
2. Reaction to task
3. Improving tasks
4. Give reward
5. Learning
6. Reaction to reward
7. Improving rewards

"POLICY UPDATING"
References