A Story of Two Streams: Reinforcement Learning Models from Human Behavior and Neuropsychiatry

Baihan Lin\textsuperscript{1,2}, Djallel BOUNEFFOUF\textsuperscript{2}, Guillermo CECCHI\textsuperscript{2}, Jenna REINEN\textsuperscript{2}, Irina RISH\textsuperscript{2,3}

\textsuperscript{1} Center for Theoretical Neuroscience, Columbia University, New York, NY 10027, USA
\textsuperscript{2} IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA
\textsuperscript{3} Mila, Université de Montréal, Montreal, Quebec H3T, Canada

Abstract

Drawing an inspiration from behavioral studies of human decision making, we propose here a more general and flexible parametric framework for reinforcement learning that extends standard Q-learning to a two-stream model for processing positive and negative rewards, and allows to incorporate a wide range of reward-processing biases – an important component of human decision making which can help us better understand a wide spectrum of multi-agent interactions in complex real-world socioeconomic systems, as well as various neuropsychiatric conditions associated with disruptions in normal reward processing. From the computational perspective, we observe that the proposed Split-QL model and its clinically inspired variants consistently outperform standard Q-Learning and SARSA methods, as well as recently proposed Double Q-Learning approaches, on simulated tasks with particular reward distributions, a real-world dataset capturing human decision-making in gambling tasks, and the Pac-Man game in a lifelong learning setting across different reward stationarities.

Split Q Learning (SQL)

Algorithm Split Q-Learning
\begin{itemize}
\item Initialize Q\textsubscript{+}, Q\textsubscript{-}, \theta\textsubscript{+}, \theta\textsubscript{-}, to all zeros
\item For each episode t do
\item \hspace{1em} Initialize state s
\item \hspace{1em} Repeat for each step of the episode t
\item \hspace{2em} Take action a\textsubscript{t} = \text{arg max}_a \{Q\textsubscript{+}(s, a) + \theta\textsubscript{+}\}
\item \hspace{2em} observe s\textsubscript{+}, r\textsubscript{+}, \text{ and } s\textsubscript{-}
\item \hspace{2em} \text{Update } Q\textsubscript{+} \text{ and } \theta\textsubscript{+} \text{ based on } (s, a, r, s\textsubscript{+})
\item \hspace{1em} \text{until } s\textsubscript{+} \text{ is the terminal state}
\item \text{End for}
\end{itemize}

Reward Processing Bias

From the perspective of evolutionary psychiatry, various mental disorders, including depression, anxiety, ADHD, addiction and even schizophrenia can be considered as “extreme points” in a continuous spectrum of behaviors and traits developed for various purposes during evolution, and somewhat less extreme versions of those traits can be actually beneficial in specific environments. Thus, modeling decision-making biases and traits associated with various disorders may actually enrich the existing computational decision-making models, leading to potentially more flexible and better-performing algorithms.

Clinical Inspirations

Table 4: Iowa Gambling Task schemes

<table>
<thead>
<tr>
<th>Decks</th>
<th>win per card</th>
<th>loss per card</th>
<th>expected value</th>
<th>scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (bad)</td>
<td>+100</td>
<td>Frequent: -150 (p0.1), -200 (p0.1), -250 (p0.1), -300 (p0.1), -350 (p0.1)</td>
<td>-25</td>
<td>1</td>
</tr>
<tr>
<td>B (bad)</td>
<td>+100</td>
<td>Infrequent: -1250 (p0.1)</td>
<td>-25</td>
<td>1</td>
</tr>
<tr>
<td>C (good)</td>
<td>+50</td>
<td>Frequent: -25 (p0.1), -50 (p0.1), -75 (p0.1), -100 (p0.3)</td>
<td>+25</td>
<td>1</td>
</tr>
<tr>
<td>D (good)</td>
<td>+50</td>
<td>Infrequent: -250 (p0.1)</td>
<td>+25</td>
<td>1</td>
</tr>
<tr>
<td>A (bad)</td>
<td>+100</td>
<td>Frequent: -150 (p0.1), -200 (p0.1), -250 (p0.1), -300 (p0.1), -350 (p0.1)</td>
<td>-25</td>
<td>2</td>
</tr>
<tr>
<td>B (bad)</td>
<td>+100</td>
<td>Infrequent: -1250 (p0.1)</td>
<td>-25</td>
<td>2</td>
</tr>
<tr>
<td>C (good)</td>
<td>+50</td>
<td>Infrequent: -50 (p0.5)</td>
<td>+25</td>
<td>2</td>
</tr>
<tr>
<td>D (good)</td>
<td>+50</td>
<td>Infrequent: -250 (p0.1)</td>
<td>+25</td>
<td>2</td>
</tr>
</tbody>
</table>

Iowa Gambling Task (IGT) with reward-biased mental agents

Ongoing directions

- Investigate the optimal reward bias parameters in a series of computer games evaluated on different criteria, for example, longest survival time vs. highest final score.
- Explore the multi-agent interactions given different reward processing biases.
- Tune and extend the proposed model to better capture observations in literature.
- Learn the parametric reward bias from actual patient data.
- Test the model on both healthy subjects and patients with specific mental conditions.
- Evaluate the merit in two-stream processing in deep Q networks.