Is Cerebellum a Model-Based Reinforcement Learning Agent?

Bharath Masetty¹,², Reuth Mirsky¹, Ashish D. Deshpande²
Michael Mauk³ and Peter Stone¹,⁴
Motivation

- Review cerebellum’s functionality from a reinforcement learning perspective
- Propose novel experiments using cerebellum simulations and ideas in RL to understand more about cognitive and motor learning in humans.
Background: Cerebellum

- Cerebellum is a major structure of the brain located near brainstem
- 10% of brain’s volume but has more neurons than the rest of the brain
- Neural substrate responsible for movement coordination and motor control
- Consists of functional subdivisions called *microzones* which modulate activity in specific muscle groups
Background: Reinforcement Learning

Markov Decision Process (S, A, p, r, γ):

- **S**: State space
- **A**: Action space
- Transition Function $P(s' | s, a)$: Probability of being in state $s'$ when taken action $a$ in state $s$.
- Reward Function $r(s, a)$: Determines reward $r$ when taken action $a$ in state $s$.
- $γ$: Discount factor

**Model-Based Methods**: Uses models of the environment to optimize the policy.
**Model-Free Methods**: Do not use models of the environment to optimize the policy.
Hypothesis

<table>
<thead>
<tr>
<th>Cerebellum Functionality</th>
<th>Process in RL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to modulate motor commands and control movement</td>
<td>Modulation of a Control Policy</td>
</tr>
<tr>
<td>Ability to predict future sensory states</td>
<td>Learning a forward dynamics model</td>
</tr>
<tr>
<td>Ability to encode external rewards</td>
<td>Learning a reward function</td>
</tr>
</tbody>
</table>

Model-based reinforcement learning *can* be one of the functionalities of cerebellum

- Literature supporting the hypothesis
- Propose a way to test the hypothesis using a cerebellum simulation
Cerebellum: Topology

- **Mossy Fibers** are the primary inputs to the cerebellum (CB)
- **Nucleus Cells** provide the primary output
- **Climbing Fibers** delivers error signals to modulate the synaptic plasticity of intermediate layers
- Divided into functional subdivisions called **microzones** controlling specific muscle groups
Cerebellum: Forward Models

- The cerebellum controls motor commands using prediction of future sensory states via its internal forward models.

- Sensory prediction error acts as a training signal to learn the internal forward models.

Hypothesis

<table>
<thead>
<tr>
<th>Cerebellum Functionality</th>
<th>Process in RL</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Ability to modulate motor commands and control movement</td>
<td>▪ Modulation of a Control Policy</td>
</tr>
<tr>
<td>● Ability to predict future sensory states</td>
<td>▪ Learning a forward dynamics model</td>
</tr>
<tr>
<td>● Ability to encode external rewards</td>
<td>▪ Learning a reward function</td>
</tr>
</tbody>
</table>

Model-based reinforcement learning can be one of the functionalities of cerebellum
Cerebellum: Reward Learning

- Recent experiments\[8\] revealed that the cerebellum has direct excitatory projections to the Ventral Tegmental Area (VTA)
- VTA also known as brain’s rewarding center
- Experiments on rodents showed that CB-VTA projects can encode external reward functions

Hypothesis

- Ability to modulate motor commands and control movement $\implies$ Modulation of a Control Policy
- Ability to predict future sensory states $\implies$ Forward dynamics of environments
- Ability to learn external reward functions $\implies$ Reward function of environments

Model-based reinforcement learning *can* be one of the functionalities of cerebellum

How to test this hypothesis?

- We propose to use a simulated cerebellum for this purpose.
Simulated Cerebellum: Related Work

- Cerebellum’s well understood topology makes it a good candidate for simulation neuroscience

- Adaptive cerebellar spiking model\cite{Niceto2011} to control robotic arm
- Cerebellum-inspired neural network\cite{Christopher2005} for state estimation and control
- Biologically constrained cerebellum\cite{Matthew2016} simulation was used to perform:
  - Cartpole balancing
  - PID Control
  - Robot Balancing using RL
  - Classification
  - Pattern Recognition

The state signal is encoded into the **mossy fibers**

- Each output comes from a distinct **microzone**
- Every microzone is associated with its own error signal
- All microzones share a common input signal
- Outputs are inferred from the firing rates of nucleus neurons
- A **pair** of microzones for each class of output
  - One of them will increase the output for that class and the other will decrease

---

Testing the Hypothesis

Model-based reinforcement learning could be one of the functionalities of cerebellum

- The goal is to test if the simulated cerebellum can act as a reinforcement learning model.

We propose a two step method:

**Step 1:** Learn the forward dynamics model and reward function of the environment

**Step 2:** Perform an n-step look ahead using the learnt model to test for policy optimization.
Step 1: Model Learning

Mossy Fiber Inputs

- \( \theta_1 \)
- \( \omega_1 \)
- \( x_1 \)
- \( u_1 \)
- \( a_1 \)

State features at time \( t \)

Action taken at time \( t \)

Microzone for \( \theta^+ \)

- Sensory Prediction Error (sPE)

- \( \theta^+_{t+1} \)

- \( \theta^-_{t+1} \)

- \( \theta^+_{t+1} - \theta^-_{t+1} - \theta_{actual} \)

- \( \theta_{actual, t+1} \)

Microzone for \( \theta^- \)

Microzone for \( r^+ \)

- Reward Prediction Error (rPE)

- \( r^+_{t+1} \)

- \( r^-_{t+1} \)

- \( r^+_{t+1} - r^-_{t+1} - r_{actual} \)

- \( r_{actual, t+1} \)

Simulated Cerebellum
Step 2: Policy Learning
Potential Outcomes

- In step 2, if the simulation is able to show improvement in its policy
  - Cerebellum simulation can perform model-based RL
  - Supports our hypothesis

- If the simulation does not show any improvement in policy
  - Hypothesis is not true, or
  - Cannot be tested within the scope of simulated cerebellum or
  - Could be due to limitations of the simulation model
Summary

● We combine popular consensus with recent evidence to hypothesize that the cerebellum can perform model based reinforcement learning

● We propose a two-stage method to test this hypothesis using a simulated cerebellum
  ○ Learn the forward dynamics and reward function of the environment
  ○ Perform an n-step look ahead on policy microzones using model microzones to evaluate policy optimization

● Potential challenges:
  ○ Biological accuracy and level of abstraction in the cerebellum simulation
  ○ Hyperparameter tuning

● Potential outcomes:
  ○ Algorithmic understanding of reinforcement learning in cerebellum
  ○ Inspiration for new sample-efficient RL methods.
Is Cerebellum a Model-Based Reinforcement Learning Agent?

Bharath Masetty
bmasetty@utexas.edu

Reuth Mirsky

Ashish D. Deshpande

Michael Mauk

Peter Stone