“Memory Transformation Enhances Reinforcement Learning in Dynamic Environments”
by Adam Santoro, Paul W. Frankland and Blake A. Richards
November 2016

Presented by Pavlo Tertychnyi
Real world example
Guide behaviour:

1. Episodic memory (short time period)
2. Schematic memory (long time period)
Agent in the environment
Simulated foraging task and the passage of time
\[ l(t) = \begin{cases} 
    l(t - 1) + \epsilon, & \text{if } t_b < B \\
    \phi, & \text{if } t_b = B 
\end{cases} \]

Here, \( \epsilon = [\epsilon_1, \epsilon_2] \), where \( \epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma_\epsilon) \). Similarly, \( \phi = [\phi_1, \phi_2] \), where \( \phi_1, \phi_2 \sim \mathcal{N}(\mu, \sigma_\phi) \).
Experiment description

The agent tasked to find 120 rewards. Performance of the last 100 were measured (the first 20 used as a pretraining to agent). Each simulation (i.e. finding of 100 rewards) counts as one sample. In all data 20 samples are present.
Basic agent architecture

Agent

Episodic memory system
Schematic memory system
Forward model for navigation
Episodic memory system

Properties:

1. Store specific reward locations
2. Emphasize more recent memories
3. Store new memories continuously
Hippocampal anatomy
Schematic memory system

It is a two-layer generative model (i.e. it stores information about probability distribution and can sample from them). The input layer is a direct copy of the place cells in the episodic memory system. Recall in the schematic system provides a new set of place cell activities modified to the expectations of the schematic memory system.
Navigation of the agent

Current CA1 Activity  Input Action

Predicted CA1 Activity

Memory Output

Action Probability
Goals for the forward model

\[ m_O = \alpha m_E + (1 - \alpha)m_S \]
Agent architecture
Comparison of memory schemes

B. Episodic ($\alpha = 1$)

C. Schematic ($\alpha = 0$)

D. Mix ($\alpha = 0.5$)
Comparison of memory schemes

- **Reward Rate (s⁻¹)**
  - Transformation
  - Episodic
  - Schematic

- **Latency to Reward (s)**

Inter-Trial Delay (s)
Shift from episodic memory to schematic
Policy computation

$$\alpha_{t+1} = \begin{cases} 
\alpha_t e^{-\beta \alpha_t}, & \text{if } R_t = 0 \\
1, & \text{if } R_t = 1 
\end{cases}$$
Dynamics of the policy unit
Habitual agent
Habitual vs memory transformation agents
Habitual vs memory transformation agents
Schematic memory in nonstationary environment

- **Pre-Training**
  - Small Gaussian
  - Reward Rate (s⁻¹)
  - Inter-Trial Delay (s)
  - Episodic (Blue) vs. Schematic (Open)

- **Early**
  - Small Gaussian
  - Reward Rate (s⁻¹)
  - Inter-Trial Delay (s)
  - Episodic (Blue) vs. Schematic (Open)

- **Late**
  - Small Gaussian
  - Reward Rate (s⁻¹)
  - Inter-Trial Delay (s)
  - Episodic (Blue) vs. Schematic (Open)
Schematic memory in nonstationary environment
Schematic memory in nonstationary environment